IEEE

MICRO

DECEMBER 1993

Chips, Systems, Software and Applications

STANDARDS
Making the pieces fit

IEEE MICRO

Published by the IEEE Computer Society

Volume 13   Number 6   **December 1993**

F   E   A   T   U   R   E   S

Cover: Jay Simpson, Design and Direction

DEPARTMENTS

*Reader Service cards, p. 96A/B;
Computer Society information,
p. 98*

From the Editor-in-Chief

Novelties coming!

BEFORE THE USUAL yearly check of the magazine's balance sheets (which will appear in the February issue), I'd like to let you know about some changes that start in this or the next issue of *IEEE Micro*.

First of all, you will find a new department (p. 95) focused on those applications of computers that significantly changed some kind of human activity. (Is there any area in which computers did not bring about changes?) We will look for unusual or unexpected aspects. Of course, your suggestions and contributions are welcome—see the editor's note on p. 96.

A second change should be most welcome to you: We've added addresses to our new products. Let me know if this helps.

Special issues have always been successful, but some readers may find a specific theme to be outside their interests. So, in 1994, we will add theme tracks. Each issue will feature one or two tracks that consist of sequences of articles on a defined topic, published in consecutive issues. This will give you wider coverage on a topic, in addition to the special issue articles.

The first track, starting in February, features Portable Computing (see announcement and call for papers on the back cover). This is a hot theme for these years, with a lot of implications: low power consumption, wireless communications, and so on. These same problems affect the big, top-performance machines. Though technologists will soon be able to put 1G transistors (10^9) on a chip, the problem of heat removal is open. Reducing heat becomes mandatory now, and this means lower power consumption too. Ideas and techniques gained from battery-powered portable systems may help us solve these problems.

A second track, planned for April, centers on fault tolerance. It will start as a continuation of the Fault Tolerant Systems Special Issue of February. The remainder of 1994 issues carry other special themes: the usual (and successful) Hot Chips (April), Analog VLSI and Neural Networks (June), Hardware-Software Codesign (August), and Optical Processing (December). (See p. 103.)

We can come to only one conclusion: Happy 1994 with *IEEE Micro*!

[signature]

Mailbag

(LK: liked; DLK: disliked; LTS: like to see)

October 1992

LK: "V42bis Standard"; "Achieving Supercomputer Performance."—F.C., Riyadh, Saudi Arabia

April 1993

LTS: Articles on communication formats.—T.D., Mountain View, CA

DLK: Publicity [is] commercial.—K.K., Mashad, Iran (But this is also a way to carry information!—D.D.C.)

LTS: CPU, graphics, hardware, interaction devices.—X.P., Geneva (Keep an eye on the next issues.—D.D.C.)

LK: This kind of special issue; LTS: more about RISC microprocessors and about the Pentium specially.—N.R. Mar del Plata, Argentina (You should also have liked the June issue then—and more is coming.—D.D.C.)

LK: Many articles in this issue, although this is the first time I've read it.... The product information is very important to me.... M.D., Irbid, Jordan

LTS: More on EDA and DSP.—C.R., London

June 1993

LK: Benchmark evaluation of chips; LTS: direct comparison of other chips. Info to guide in conversion from proprietary minicomputer to Unix (open) machines.—D.F., Euless, TX

LK: Emphasis on architectural features in the microprocessor articles; DLK: "Alpha AXP" did not clearly explain why Alpha is scalable.—B.M.K., Toronto, Canada (I forwarded this question to the authors.—D.D.C.)

LK: Micro Law.—P.W., Morris Plains, NJ (I'm in full agreement! It is one of our strongest points.—D.D.C.)

LTS: More information on testing equipment for various kinds of chips in the New Products section.—M.G.H., Rehovot, Israel

August 1993

LK: "Dream Chip"—excellent concept; LTS: article on RISC MIPS; what a CISC equivalent would be; memory subsystem architectures.—E.B., El Toro, CA (The new department for "dream" chips is open to readers. Would you like to contribute your idea?—D.D.C.)

Micro News

Send information for inclusion in Micro News one month before cover date to Managing Editor, IEEE Micro, PO Box 3014, Los Alamitos, CA 90720-1264.

NSF funds all-optical network research

Seen as the backbone of the nation's future communications systems is an all-optical network that researchers feel will transmit information 10,000 times faster than the current few billion-bps transmissions. Though today's patchwork of optics and electronics accommodates the Clinton administration's plans for a national data super-highway, the all-optical approach would become the logical next step as information exchange demands continue to increase.

To develop the links for US telecommunications and computing industries, the National Science Foundation this September awarded a $1.89-million grant to a research team from Stanford University, UC Berkeley, and UC Santa Cruz. The team will 1) pursue the idea of "tuning" an optical fiber to pick out the wavelength of interest from different wavelengths and 2) create switches that route signals based entirely on optical information. Each packet of information—a telephone call, an e-mail message—will carry with it a unique tag that identifies it and tells each switch along the way where to send it.

—UC Newswire

Information highway concerns

Computer Professionals for Social Responsibility, a California nonprofit, public interest organization, warned this October that the planned information superhighway will not realize its full potential. CPSR president Eric Roberts urged the Clinton administration to quickly adopt specific goals to ensure the National Information Infrastructure is used as a public resource.

According to CPSR President and Associate Chair of Stanford's Computer Science Department Eric Roberts, "... a small number of companies [could come] to dominate the market, [placing us] in danger of stifling competition and innovation on the network. If those same companies control the programming, then open and diverse speech is limited. If pricing structures do not cover universal service, the average person and the poor will be struggling to use the backroads of the information highway." Other concerns centered on privacy rights and equitable and universal access.

CPSR's NII paper is available electronically at listserv@cpsr.org and on hard copy by telephoning (415) 322-3778.

Standards galore!

Recent developments in standards projects include activities for LANs, SCI products, Unix, and Futurebus+. Earlier projects included SBus I/O bus acceptance by the IEEE as Std. 1496-1993 plus start-ups for high-performance token architectures and P1355 for Heterogeneous InterConnect, which resulted from ESPRIT Project 7252. For information on these earlier announcements, telephone the Computer Society at (202) 371-0101.

LAN projects. The IEEE authorized two separate access method standards projects for 100-Mbps LANs this September. P802.12, a demand priority access method, and P802.30, a CSMA/CD project, were proposed by the LAN MAN Standards Committee of the Computer Society. For more information on P802.12, telephone Pat Thaler, Hewlett-Packard Co., (916) 785-4538; for P802.30 information, contact Peter Tarrant, SynOptics Communications, Inc., (408) 764-1217.

P1596.x, SCI implementations. Soon to accompany SCI GaAs chips available from Vitesse Semiconductor are CMOS chips from LSI Logic and Dolphin, which will interface directly to the HP GLink chips for serial and fiber-optic links at 1.25 Gbps. Further implementations are expected from Unisys and Convex Computer.

Unisys will build 1-Gbyte/s SCI chips for early 1994 release. Convex Computer plans to use SCI

as an "enabling technology" for its next-generation of shared-memory multiprocessor supercomputers. Copies of the standard sell for $68 to non-IEEE members and less to members.

Unix standardization. Novell Inc. has agreed with X/Open Co. Ltd. to establish a single Unix specification backed by a Unix trademark. Under the agreement, Novell transfers ownership of the trademark to X/Open, the international open systems standards organization. Products bearing the Unix trademark will adhere to the Spec 1170 application programming interface specification, which allows choice of open systems vendors and improves network integration.

Futurebus+. Promising to uncover flaws in this IEEE standard is a Carnegie Mellon-developed tool more exact than simulation. Model checking uses temporal logic formulas and Boolean decision diagrams to search all possible states of a chip up to 10^{100}. Intel uses model checking as a chip production tool; it is under investigation by Hewlett-Packard, Bull S.A., Siemens, Fujitsu, and several universities.

From gooey to sticky?

Personal Digital Assistant interfaces are changing. Apple, which commercialized the GUI (pronounced "gooey") graphical user interface, uses the STICI (pronounced "sticky") Self-Teaching Interpretive Communicating Interface in its Newton MessagePad.

First used with Go's PenPoint, STICI adds Newton Intelligence to let the operating system and interface study how a user makes use of a PDA, make inferences based on that data, manage PC/PDA communications functions, and interact with documents.

Plans call for seamless linking of PDA tools from multiple vendors for all user-written or -drawn documents. This openness has led BIS Strategic Decisions in Norwell, Mass., to predict that 60.4 percent of US PCs and PDAs will use STICI by 1998.

Digital HDTV progresses

Technology decisions on key building blocks to make up the digital high-definition television system being proposed to the US Federal Communications Commission have been announced by the "Grand Alliance." Digital video compression, transport, scanning formats, and audio technology were endorsed by the FCC's Advisory Committee on Advanced Television Service. These decisions incorporate modifications of the GA system that had been recommended earlier by the Technical Subgroup.

After full Advisory Committee approval, the GA can proceed with construction of most aspects of the prototype system, which is expected to be tested in 1994. In early 1994, following competitive testing of the 4-VSB, 6-VSB, and 32-GAM broadcast modulation systems and related cable modes, the alliance plans to select the broadcast and cable transmission technology.

This alliance represents the merging of technologies developed by the three groups that vied for the digital HDTV standard in the US: AT&T and Zenith, General Instrument and MIT, and the Thomson, Philips, and David Sarnoff Research Center consortium.

High-speed recognition chips

A November agreement between Nestor, Inc. and Intel Corporation set the terms for commercialization of their jointly developed Ni1000 neural network chip. By combining the Ni1000 with the NestorReader OCR, the companies intend to offer a line of high-speed (output in pages/s, not characters/s) optical character recognition products in 1Q94.

Launched in 1990 with funding from DARPA, the Office of Naval Research, Intel, and Nestor, the Ni1000 project supplied beta sample chips, development tools, and technical support to system developers. A later ARPA contract funded subsystems.

For the record

IEEE Micro's October issue carried a typo in Figure 10 on p. 63. The lower right-hand corner of the PowerPC memory path block diagram should have read "64-bit" not "84-bit."

Reader Interest Survey

Indicate your interest in this department by circling the appropriate number on the Reader Service Card.

Low 198   Medium 199   High 200

On the Edge

Developing 3D memories
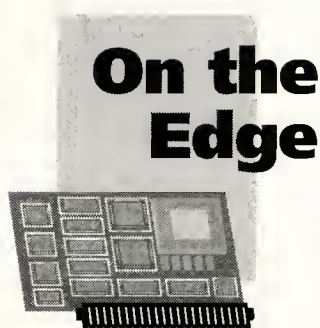
Arun K. Sood

George Mason
University

Fairfax, Virginia 22030

asood@cs.gmu.edu

[On The Edge invites contributions of short tutorials on cutting-edge topics in hardware, microsystems, and software. We encourage short notes on emerging applications, markets, and "dream" chips (see an example in the August 1993 issue, p. 49).

Publications for On The Edge are reviewed on a fast track. Also, we welcome suggestions from readers identifying topics on which they would like to see short tutorials.

Here, John C. Carson, Myles F. Suer, and Raphael R. Some of Irvine Sensors Corporation discuss their work on 3D stacking technology. After first funding by DARPA for sensor applications and computer memory packaging, additional agencies provided support for memory short stack technology. This enthusiastic response convinced the company and IBM Corp. to jointly develop commercial products.—A.K.S.]

Three-dimensional IC memory stack technology enables a new generation of lower cost, higher performance, and higher reliability computer electronics. Although 3D packaging is only beginning to emerge from the laboratory, its obvious cost/performance and cost/benefit characteristics will allow insertion of the technology into both high- and low-end systems.

Two types of memory chip stacking technologies—full stack and short stack—are in the final stages of development now.

The full stack places up to 100 ICs in a "loaf-of-bread" configuration. A typical full-stack size is $2.5 \times 1.25 \times 0.65$ cm. The short stack puts 4 to 16 ICs in a "stack-of-pancakes" configuration. The resultant stack resembles a thick IC with typical dimensions of $1.25 \times 0.65 \times 0.025$ cm. The two stack form factors support computer memory applications. While the full stack typically interconnects to the next level of assembly by bump bonding to a substrate or wire bonding into a deep-drawn custom package, the short stack works with wirebonding or tape-automated bonding into standard packages, SMT, or chip-on-board mounting.

Evolution of full stacks

Stacking technology has evolved to the level that we can directly laminate thin ICs and bring die interconnections to one or more faces of the resulting "cube." This also provides IC lead interconnection and busing on the cube itself. We first applied this technology to custom, cryogenic CMOS ICs used in infrared sensor signal processing. Subsequently, we adapted to GaAs and CMOS memory under funding from the US Defense Advanced Research Projects Agency (DARPA). The process first reroutes the I/O connections of standard ICs at the wafer level. After lamination, we metallize the cube face—with address, control, and data I/O—with individual pads. The cube is then ready for "bumping" to a substrate or assembly in a custom package.

The alternatives for full-stack packaging include single- and multicube packages using high bonding shelves and/or superstrates for wirebonding or TAB interconnection of the cubes to the package leads. As we obtain volume production levels, the per die cost of full-stack packaging should reach parity with plastic (TSOP) packaging. The full stack provides the following features:

- reduced volume, line capacitance and drive requirements,
- lower number of bypass capacitors,
- elimination of memory bus line resistive termination networks,
- elimination of most of the drivers and decoders required to support large memory systems,
- reduced system noise by using low-power drivers,
- reduced signal propagation delay and required settling times,
- improved system reliability, and
- good heat dissipation, typically resulting in a 1°C temperature rise per watt.

Once memory developers design die with the 3D IC stack in mind, improvements in performance, reliability, and cost will provide system-level enhancements normally associated with generational advances in IC technology.

Evolution of short stacks

Development of short stacks occurred for two reasons: the inability of many potential customers to absorb the "head room" of the custom packages required by the full stack, and the desire to provide a component with the benefits of 3D stacks that handles like a standard IC. The top layer is a ceramic "cap chip," which provides bond pads suitable for wirebond, Tab, SMT, or chip-on-board (COB) technologies. Short-stack development was initially sponsored by the USAF Rome Laboratories and subsequently by NASA.

Full-stack applications

We have defined several applications for full-stack memory components. Candidates for near-term commercial production include main memory for mainframe computers, cluster memory for parallel processors, and cache memory for high-performance disk systems. All these applications feature similar full-stack component architectures and DRAM.

Mainframe main memory. Large mainframe computers require massive amounts of DRAM. The memory subsystems are often the largest, most costly, and most unreliable segments in these machines. The use of 3D memory full-stack components allows the reduction of the subsystem from 10 or more boards to a single board. In addition, we reduce the number of drivers, decoders, and memory board controllers. System bus loading is similarly lower, leading to reduced bus driver requirements. The reduction in system size allows an improvement in memory access times and a general increase in system bus speed. The reduction in interconnections signifi-cantly improves system reliability. These improvements enhance performance with no attendant increase in cost.

It is also possible to use the density, cost, reliability, and performance benefits of full stacks to provide additional main memory in the basic system. Additional main memory allows more of the system and application software to remain resident, thus minimizing software access delays and enhancing system-level performance. Using 100-layer full stacks provides an improvement in packaging technology for main memory broads of more than 50 to 1.

Multiprocessor cluster memory. Many current multiprocessor architectures revolve around the concept of a cluster node. A cluster usually consists of four processors sharing a main memory segment and communicating with other clusters via high-speed busing. These architectures promise many of the benefits of both shared-memory and distributed-memory systems by providing elements of both in a single unified system. To maintain high performance, to keep costs reasonable, and to keep the system size within manageable limits, designers prefer to place a cluster on a single board. To do this, they need advanced packaging of the shared cluster memory. And, 3D DRAM full stacks provide a simple and effective way to do this. In a typical cluster board comprising four pro-

> *These characteristics will allow insertion of the technology into both high- and low-end systems.*

cessing elements (PEs), a full-stack 3D DRAM, and an external network interface, not only is the physical size of the memory significant but also its speed, since it must service five ports (4 PEs + network I/O). The full-stack implementation provides the enhancement required to attain maximum performance from the shared cluster memory. In addition, like the mainframe memory, reliability, system complexity, and power utilization are significant system discriminants. Full-stack technology provides the required improvements in these areas over conventionally packaged systems.

Disk cache

The search for more powerful and responsive bulk storage has resulted in the inclusion of cache memories in high-performance disk systems. To be effective, these disk caches require dense, cost-effective IC packaging. A full-stack implementation fits well within the envelope of the disk housing or on the disk controller board, is extremely reliable, and (assuming 16-Mbyte DRAMs) will provide over 200 Mbytes of on-line, high-speed data buffering.

Short-stack applications

We have also defined several applications for short-stack memories. Three

Guest Editor's Introduction:
Reengineering Standards

Stephen L. Diamond

Picosoft, Inc.

When standards define an interface—that infinitesimally thin boundary layer across which two disparate worlds can communicate—they facilitate progress by focusing our attention where we can have the most positive impact.

Standards in the best sense are interfaces, not just between networks, or computers, or chips, but between people and organizations and ideas. In my view, the purpose of standards is successful products—not standards per se. But the discipline of standardization must reengineer itself if it is to meet today's challenges.

Bring up standards in polite company and you may see your listener's eyes instantly glaze over, or find yourself in the midst of a debate on how standards obstruct innovation. But ask your listeners about the impact of the VHS standard on consumer electronics, and you may get an entirely different response. Why? Because many people think of standards as only academic and obscure documents full of arcane language produced at glacial speed by near-retirees in century-old organizations.

Today's standards are agreements, not just documents. These agreements are the critical enabling technology, the paradigm shift that will change new industries, like the nascent multimedia information highway, or wireless communications, or personal digital assistants. These agreements can be produced in different ways, including emerging electronic standards development methodologies being explored by organizations such as the IEEE Computer Society, as well as Information Technology (IT) consortia. The ramifications of alternative standards development strategies are complex, with far-reaching consequences.

Standard terminology

An interface is a defined multidimensional surface at which independent systems interact. Think of the interface between humans and automobiles as just the accelerator, brake pedal, and steering wheel. This interface defines how humans interact with the automobile; it does not define how stepping on the brake causes the car to slow. The brake pedal may be connected mechanically to drum brakes or hydraulically to disc brakes, or it may cause a parachute to deploy. They all implement the same interface.

That is the power of interface standards: They encourage multiple, innovative implementations to compete and evolve in the marketplace. By combining a sufficient number of these interface standards, one can define all the elements of a computer—a virtual machine—that can be implemented in many ways, by many companies, and at many levels of cost and performance.

In contrast, a standard based on one particular implementation discourages innovation, and might, for example, have required us to drive cars with mechanically operated nonpower-assisted drum brakes. The final arbiter of the value of a standard should be the marketplace, not a standards developing organization (SDO). So long as the appropriate process is followed, and users and vendors are willing to develop a standard, anything is an appropriate topic for standardization.

We have two major categories of standards, *de facto* and *de jure*. Accredited SDOs give us de jure (according to law) standards. The American National Standards Institute sets the rules, or metastandards, for US SDOs, which require openness, consensus, and due process. ANSI is the US member of ISO, the International Standards Organization. For de jure standards in the IT industry, ISO standardization is the optimum goal—a single standard accepted all over the world.

Not surprisingly, there are no agreed rules defining what is a de facto (in fact) standard. Market acceptance is often thought of as a synonym, although that too is an intuitive metric. No rules

of openness, consensus, or due process apply, and a single company controls many of these standards. Of course, de jure standards may evolve into de facto standards, allowing the industry to enjoy the benefits of an open process and the economies of scale created by a volume market. The converse (de facto standards becoming de jure) is also possible and desirable. Some user groups and consortia seek to foster de facto standards in a formal, structured manner. Such organizations have rules that define the process and may seek to involve interested constituencies. But because they answer to no accrediting authority, they do not develop into de jure standards.

One final term I want to introduce is *standardization*—an engineering and marketing discipline that seeks to foster successful products through the development of de facto and de jure standards. Standardization offers an unusual opportunity to advance a company's products, and at the same time, benefit the industry and the profession.

Pragmatics

Companies participate in standardization because they perceive an economic benefit. Shane Greenstein discusses some of the economic implications of decentralized mechanisms of standardization, with regard to what are called economic networks—all the buyers and sellers economically interested in the features of a system.

Should government become more involved? Linda Garcia discusses the history of public and private sector roles in standards development, and raises the possibility of greater government involvement.

Choosing a standards strategy. The critical questions are: Where are we today in standards? Where could we be tomorrow? Where *should* we be? Do we want an open standard or a proprietary product? Table 1 lists four strategies of standardization (leading, following, monitoring, and ignoring) and their characteristics.

Leading requires a large investment and carries with it a great deal of risk: There may be no followers, or another interface may compete for leadership and win, necessitating a retrenchment. Following involves a lower cost strategy, and several organizations may cooperate as leaders and followers to create a camp, as Chris Halliwell discusses. Monitoring is a passive strategy, and accordingly, the risk is greater because the leader is less constrained from changing direction. Companies that believe they are ignoring standards are really implementing a leading strategy. But by misunderstanding their strategy, they lose the opportunity to use explicit standardization techniques to promote their interfaces.

Standardization options. Once chosen, the options involve choosing avenues for developing and promulgating the standard: The basic choice is either open or closed. Either approach can lead to the primary goal of a de facto standard. The de jure route is more difficult, but is often the

Table 1. Standards strategies.				
Strategy	Cost	Risk	Potential for cooperation	Time to market
Leading	High	High	Low	Short
Following	Medium	Low	High	Medium
Monitoring	Low	Medium	Medium	Long
Ignoring	High	High	Low	Short

only way the industry will accept the interface. The IEEE is a major developer of de jure standards, with more than 600 currently active and over 700 in development.

New technology and new products frequently demand new interfaces. Litigation today has become a major consideration in development and adoption of new standards. Gervaise Davis discusses the inherent conflicts between US law, compatibility, and standards.

Two fast-moving and controversial applications of standardization are in the area of Asynchronous Transfer Mode (ATM) and data encryption. Richard Vickers addresses the parallel development of ATM technology and standards. Burt Kaliski provides us with a road map of the variety of encryption algorithms, and the multiple and competing standards definitions that support them.

SUCCESSFUL STANDARDIZATION REQUIRES an unusually broad range of disciplines (all with a global perspective): engineering, marketing, economics, group dynamics, law, and finance. Individuals and companies must also understand the history of standards, and the processes, precedents, people, and organizations.

SDOs, consortia, and user groups are reengineering their processes to make them more responsive, comprehensible, and agile. Change must come if the discipline is to survive and achieve its potential. I hope this special issue will help you achieve a broader perspective on the purpose, use, process, and future of standardization. I believe this perspective will help you in achieving the real goal of standards: successful products.

Stephen L. Diamond, guest editor of this special issue, is president of Picosoft, Inc., a high-tech consultancy in marketing, strategic planning, and standards. He is also the chair of the CS Microprocessor and Microcomputer Standards Committee. Direct questions about this special issue to him at Picosoft, Inc., PO Box 5032, San Mateo, CA 94402; diamond@picosoft.com.

Camp Development:

The Art of Building a Market Through Standards

To reach the volume market for a new technology quickly, and to encourage widespread industry deployment, standards development participants must stabilize standards by enlisting the support of a synergistic group of related solution suppliers called a camp. As camp members deliver a coherent and complete solution, user costs and risks are reduced, thus accelerating new technology adoption.

Chris Halliwell

Although the electronics community has made great strides in shrinking time-to-market cycles, much less progress has been made in reducing time-to-mainstream-market acceptance of new products. Technology products generate substantial profits when they garner the volume of this mainstream market, and one critical ingredient for acceptance of new products in the mainstream market is the development of standards. The primary economic value of both formal and informal standards is in lowering the buyer's total cost and risk of a new product purchase, thus accelerating adoption of the new technology. This raises the question: Are standards actually being developed in such a way that the customer's risk of acceptance is reduced?

Understanding a standards marketing technique termed *camp development* can help us answer this question. Camp development is a way of mobilizing a group of market participants around a standard so that accelerated adoption of new technology is encouraged by lowering the customer's cost and risk. Although the concept of camp is the essence of driving a market standard, it may or may not involve formal standards-developing organization approval. Standards development participants in both formal and informal organizations can apply the principles described here.

Camp development methodology is founded on an understanding of how markets develop by meeting user needs. We must understand the needs, concerns, and roles of new technology users. For example, early market customers validate and help establish standards, while mainstream users adopt standards once they are stable. Also of importance is the relationship between the standards development and market development processes. I briefly review four standards efforts, define four main types of camp participants, and illustrate camp development techniques to help fill in the details.

A final point to consider is the implications of camp development for the formal standards development process. The purpose of standards bodies is to assure technical quality and interoperability among products from multiple industry participants. With technology advancing and new markets evolving at such a rapid pace, standards organizations must respond swiftly and increase their understanding of the role they play in market development. Camp development thinking encourages formal standards committees to assure that their actions contribute to accelerated market development.

Early markets

In the 1950s a technology diffusion model emerged from a study of the way the telephone was adopted in post-World War II US suburbs. That model evolved into the adoption cycle,

Mainstream

	Innovators	Early adopters	Early majority	Late majority	Laggards
Character	Technology for its own sake	Technology for competitive advantage	Practical technology for productivity	Want technology to be transparent	Technology-averse
Standards orientation	Will connect incompatible components	Will provide integration support	Want stable, application-specific standards	Demand standards-based, complete packaged solutions	Cling to old standards
Role	Verify the underlying technology	Pragmatic deployment of proposed standards	Their demand justifies industry investment in standard	Widespread standards adoption	Extend life of previous technology

Figure 1. Standards and the technology adoption cycle.

which is often applied to new technologies and new products. This model, seen in Figure 1, forms the basis of the following market descriptions. The fundamental insight for technology marketers is that adoption is largely a function of the buyer's willingness to accept the level of risk embodied in a new technology or product. While not all new products incorporate revolutionary technology, I use the phrases "new product" and "new technology" interchangeably here.

This cycle traditionally refers to the first 2 to 3 percent of new technology buyers as innovators. This group doesn't need standards because they typically buy one of each new technology product on the market. Innovators are characterized as technologically savvy tinkerers who will connect incompatible components if necessary just to try them out. They are willing to take the risk involved in adopting new technology because they will find a solution to most technical problems. This type of buyer purchased the first personal computer, the MITS Altair, before operating systems or disk drives were available. But, even though they want to hang on the leading edge, they are not going to deploy the technology widely.

The innovator's primary buying motivation is the promotion of technology for its own sake. Regis McKenna, the Silicon Valley marketeer, recounts a venture capital community joke that new computer company business plans always boast closure of "our first seven customers." The joke is that this is always the same seven customers: JPL, Lawrence Livermore, NASA Ames, and the like. This buyer values access to all vendor technical data. Its primary roles in market development are to shake the bugs out of the new technology and to validate that it basically works as advertised. Examples of standards that have recently achieved acceptance from this category of buyer or stage of market development include those for full-motion video compression, high-definition television, and many high-bandwidth data-transmission protocols.

Following in the cycle after the innovator, the early adopter represents up to 13 percent of the market. These buyers are the visionaries that see the business value of new technology and are therefore willing to take the risk. However, they don't need standards either, for two reasons. First, this buyer is primarily motivated to achieve a competitive advantage and therefore wants to implement the technology before its more pedestrian competitors. Second, this buyer is typical-

> ***Consider US government estimates that for every dollar spent on new hardware, $10 is spent on software, $100 on data, and $1,000 on training.***

ly systems savvy, possessing the systems integration skills and resources necessary to cover the cost of achieving interoperability. Experienced electronics and computer marketing people know these companies: Apple, Sun Microsystems, American Airlines, Merrill Lynch, Hyatt Hotels, Federal Express, and others.

Early adopter companies talk to each other across industry segment boundaries about the application of new technology. They keep tabs on the innovators to see what is new on the horizon. They value quick time-to-market and are therefore very sensitive to project milestones—to getting things done on time. They serve important functions in market development.

First, they can provide the context for discussion in the market about the business value of new products, not just technological merit. Second, because they add utility to new products upon integrating them into their current environment, they can lead new technology suppliers to identify all the necessary elements of a complete mainstream customer solution—things like training classes, integration services, support and development tools, system software and drivers, and the like. Lastly, these early customers make real investments in new technology, and pragmatically implement proposed standards. As such, they provide the beginnings of a testimonial reference base necessary for more risk-averse market participants, including standards bodies, to feel comfortable adopting the new technology. Examples of standards that have achieved acceptance from early adopter companies are asynchronous transfer mode (ATM) by Sun Microsystems and digital cellular implementations in Europe by Ericcson.

Mainstream markets

This group of customers, traditionally known as the early majority, can represent up to 30 percent of the market for a new technology or product. These pragmatists are significantly more risk-averse than early buyers. They are the practical business people who value incremental productivity gains and investment protection. Whereas the early adopter

is acutely aware of the total business benefit enabled by the new technology, the pragmatist is equally sensitive to the total cost of adoption of new technologies. These costs can be substantial. Consider US government estimates that for every dollar spent on new hardware, $10 is spent on software, $100 on data, and $1,000 on training.

Mainstream buyers possess neither the resources nor the inclination to create a total application solution around a new technology. They expect the product to be packaged and specific to the application. Often, new technology vendors either don't plan for or can't afford the investments necessary to lower adoption cost. New technologies that neglect to address the mainstream customer's adoption costs fail to ever achieve a volume market despite unit price reductions. The product sinks into a "chasm" of slow sales growth, and is ultimately overtaken by alternative technology. Some famous examples of chasm products include bubble memories, integrated services digital network (ISDN), and artificial intelligence.

Accepted standards are critical to the mainstream market customer's decision to adopt new technology. For instance, both formal (de jure) and informal (de facto) standards reduce adoption cost elements for personal computers. Informal, de facto standards govern the interface between processors, operating systems, and application software to lower the cost of integration. Hardware and software component interfaces and networking standards act to lower the cost of connection. Standards such as file, image, and document formats lower the cost of operation. Formal standards body (called SDO for standards developing organization) approval of interfaces further reduces risk by assuring investment protection: that the specification is technically sound, will remain available, and will not change at the whim of one or a few vendors.

One other important characteristic of the early majority affects the standards development process. This mainstream market is segmented and specific to application. The complete customer solution—the services, the software, the education, the answers to risk questions, and the interfaces—differs from application to application. Often, earnest disagreements between market players are finally resolved when the standards process acknowledges a segment difference. For instance, consider the mainframe application orientation of token-ring networking versus the distributed computing model of Ethernet, or the various bandwidth implementations of Ethernet itself. Other examples include industry-specific standards such as those for barcoding or electronic data interchange. SDOs and informal standards groups must be quick to identify and accommodate segment standards to encourage mainstream market development.

Standards and market development

Despite the power of standards to encourage buyer and

supplier investment in new technology, the formal standards-setting process has not kept pace with the rate of technology change, or of new product development cycles.

We still measure the standards process in multiples of years. For instance, it took nearly six years to achieve official IEEE approval of Multibus I in 1982. It continues to take three to five years for CCITT adoption of new modem standards—about the same amount of time it takes modem makers to drive the market to the next bandwidth. The SCSI-2 standards process begun in the 1980s was never completed. Standards marketing people have long understood that participants in SDOs are strongly influenced by the actual installed user base when deciding on interface definition. Many market development programs and the various types of informal standards-setting techniques employed by suppliers today originally evolved from the need to create enough of an installed base to sway or accelerate the formal standards development process.

A look at the history of Multibus standards illustrates the basic principles of how the standards development process relates to market development. Multibus began as an Intel proprietary bus technology for a relatively small circuit board market. Intel realized that its financial success required a broad market for single-board computers and peripherals, and that it could not itself provide the full range of board technology to users. To grow the market, and to differentiate itself from Digital Equipment Corp., Intel polished its specification and took it to the IEEE Microprocessor Standards Committee in 1977. As the specification languished in subcommittee, Intel increased its marketing efforts by forming a manufacturer's group. The IEEE formally anointed Multibus technology as IEEE Std. 796 in 1982.

The important principle at work here for both standards development and market development is that Intel very effectively enlisted the support of its technology partners—its competitors really—toward a common goal of market growth. As this example shows, there is not much difference in achievement between approval from an SDO and the efforts of industry consortia or even of a few dominant market players. While the formal process offers technical consensus and openness, an informal approval may prove swifter and more flexible. As one Multibus standards veteran says, "You can't use standardization as a substitute strategy for a good product and good market development, but you can use good market development programs as a substitute for standardization." This statement emphasizes the importance of creating guidelines for market development—later described as camp development—that make all types of standards-setting efforts more effective.

Early Ethernet market development and standardization activities took a similar but even more effective approach when Bob Metcalfe drafted a letter to Xerox executives suggesting a joint effort with Digital. He made the argument that

Many of today's standard-setting techniques evolved from the need to create enough of an installed base to sway or accelerate the formal standards development process.

where office automation and computers were concerned, compatibility was the basis for user value.[1] He suggested that both companies should worry about market size, not market share. Next, Metcalfe brought Intel into the core group to add high-speed chip expertise. This synergistic trio, which now included a systems manufacturer, an office systems and peripherals vendor, and a semiconductor supplier, did not see each other as direct competitors. They were, together, capable of providing the basic hardware elements of a more complete end-user solution. While the IEEE Std. 802 committee pondered the specification, early adopter minicomputer vendors were signed up to support it, and these agreements were actively marketed by the triumvirate in ads and public relations. The IEEE approved the standard in 1983 after three years of consideration.

Contrast this effort with the more recent, doomed standards campaign of the Advanced Computer Environment (ACE) initiative. Here, competing microprocessor vendors, competing systems vendors, and competing operating system vendors failed to achieve a common strategic agenda on market growth. Both the standards-setting effort and the market development effort to achieve a so-called open architecture degenerated into a confusing array of camp battles and conflicting goals.

Another recent example further illustrates the underlying principles of standards and market development: the Micro Channel Architecture (MCA) and Extended Industry Standard Architecture (EISA) personal computer bus wars of the late 1980s. IBM announced Micro Channel as a "standard" in 1987, amid much expensive hoopla, but users expressed confusion about its benefits. EISA followed two years later with a better accepted specification, but no early market adopters. Neither IBM nor the Compaq-led EISA camp appear to have been trying to increase the PC market via standardization; rather, both efforts seem aimed at market control and share interests.

Clearly, the success of these two would-be standards did

Market development principles as they apply to standards

1. Suppliers participating in the standards process must focus on a common strategic goal of growing the market.

2. Standards developers must cultivate early users who make investments in new technology and deploy proposed standards.

3. Mainstream market growth requires assembly of a set of partners pledged to support a standard and capable of delivering an end-user solution.

4. To attract more risk-averse mainstream buyers, and to lower adoption costs, market solutions and standards themselves must be application specific.

5. All standards participants must invest in market development and education programs.

not hinge solely on the technical merits of the specifications. By November of 1990, Aaron Goldberg at *Computer Reseller News* wrote that, "The user demand for the advanced features of MCA and EISA has been minimal."[2] Nor did success spring from the number and clout of cooperating competitors who signed up—EISA's Gang of Nine and, later, the Micro Channel Developer's Association. Goldberg and others reported infighting in both camps and disloyalty among players who covered their bets by developing both buses.

Neither MCA and certainly not EISA achieved the level of industry acceptance necessary to be considered a true standard. However, one bus became a mainstream technology because it found a home as part of a complete end-user solution. MCA did reach a reasonable volume market because IBM included it in a well-thought-out, well-packaged, safe product called PS/2. EISA, on the other hand, was included in the Compaq Systempro, which has not garnered much mainstream market acceptance because it lacks the multiprocessing software and peripherals necessary for a complete server application solution.

The EISA/MCA story also illustrates the basic principle that the mainstream market is highly segmented by application. In retrospect, the higher performance, higher cost EISA was aimed largely at the server market, while the slower, lower cost MCA found a home in the traditional segments of IBM solution strengths—point-of-sale and data-entry applications. Had the segment focus been acknowledged at the time, each set of standards advocates could have focused on enabling complete application solutions, rather than focusing on com-

peting with each other.

The adjacent box summarizes the important market development principles that should be applied to standards development processes.

Camp development

This market development methodology lets participants in both formal and informal standards-setting efforts accelerate technology adoption. A synergistic camp that both minimizes competitive issues and aids each company in pursuing its own economic best interests requires four primary types of members. The first are the technology partner companies with interdependent technologies that absolutely must support the standard—second sources, or a microprocessor vendor and an operating system supplier, or Apple and Adobe in desktop publishing.

Influential users who actually implement the standard make up the second set of camp members. These early customers, while not the mainstream market, are extremely influential and visionary organizations. They provide informal blessing and help guide the technology partners in defining all the elements of a complete market solution. The best-known industry example is IBM's adoption of the x86 architecture and DOS. IBM's use of these technologies gave potential mainstream users a sense of safety, and the hint of mainstream demand attracted Compaq, and then ultimately hundreds of other suppliers to the standard.

The third type of camp member comprises the largest subset of the group. These innovative companies represent each piece of the customer solution set. For instance, the new technology may be aimed at OEMs that have design issues that necessitate having development tool vendors in the camp, or suppliers that make related peripheral technologies. This customer may not have the software to support products based on the new standard, so algorithm or application software vendors may also be included in the camp. The customer may not have the service and support channels needed to sell the end product enabled by the new standard, so consulting, training, and distribution partners may be recruited into the camp.

This elaborate set of players is entirely necessary because each relies on the others to provide all the elements of end-user value, and to jointly invest in market development for the standard. Some of these partnerships are technically independent of the standard specification; in other words, they may not have a physical interface to the hardware or software standard. Each partner does, however, make a learning investment in the standard, and stands to profit by its acceptance.

The fourth type of camp member is comprised of gurus and market pundits who influence market development by communicating evidence of the progress that the camp makes toward providing mainstream market solutions. This is an Esther Dyson in PCs, or a Forrester Group in network-

ing, or a Michael Slater in microprocessors, or a Carver Mead in semiconductor design methodology, to name a few. This group also includes other organizations that explicitly or implicitly sanction the standard, including benchmarking and testing labs, educational institutions, government agencies, associations, and others. These members not only build end-market demand for new technology by creating an impression of "safety," they also have an indirect effect on the standards development process. They publicly verify that a mainstream market will emerge around a particular standard.

Establishing and maintaining camp relationships is a time-consuming process that requires management involvement and dedicated resources. The selection of camp members should favor a few partners—one or two of each type, and a complete solution set of players—ready to make real technology and market development investments. Ideally, technology partners form the first relationships in the camp. Next, the technology partners seek early customers, and then solution partners. With that accomplished, they look for influencers to join them. Each relationship is both encouraged by, and reinforces, prior relationships. The net effect is what we call market "momentum" (see Figure 2).

AT&T executed this kind of key partner recruiting during its first year of informal standard and personal communicator market development activities for its Hobbit microprocessor. The AT&T effort is not an example of perfection, or even of ultimate success since it is too early to tell, but of advanced camp development methodology. Whether AT&T will reach its market goals, or whether its camp standard will prevail, is not the subject of this article. In fact, one of AT&T's chief competitors in this new market, Apple, has been practicing forms of camp development in the personal computer market for many years. I use AT&T as an example simply because I know something about the AT&T personal communicator program through prior efforts leading a team of consultants at Regis McKenna, Inc. The list of 1993 Hobbit personal communicator camp members appears in Figure 3.

Technology platform partner programs

AT&T's decision to market the Hobbit microprocessor was founded on its relationship with Go Corporation and on Hobbit support from the Penpoint operating system. This relationship was significant because software typically drives new microprocessor adoption and the informal standards-setting process. Forming a strategic technical relationship to drive a market standard and create a new market is a significant challenge for management, requiring personal executive commitment and experience. Companies that are successful in jointly driving the market possess a number of important qualities, such as open-
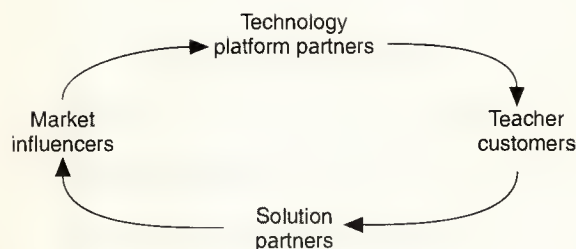

Figure 2. How a camp creates market momentum.

ness, trust based on reliability, and the ability to break down cultural barriers between the organizations. The strategic partners must implement a well-defined process for cross-organizational learning and program execution, and like the AT&T/Go team, implement a definition of joint strategic market goals.

Often, technology partner relationships are not exclusive for either side. Both participants must be able to articulate the strategic value of the relationship beyond straight financial gain. Both must be able to answer questions like Why this partner? or How is this partnership different from others?

Since the first implementation of Penpoint was for another microprocessor than Hobbit, it was critical that AT&T and Go work out answers to these questions. The successful
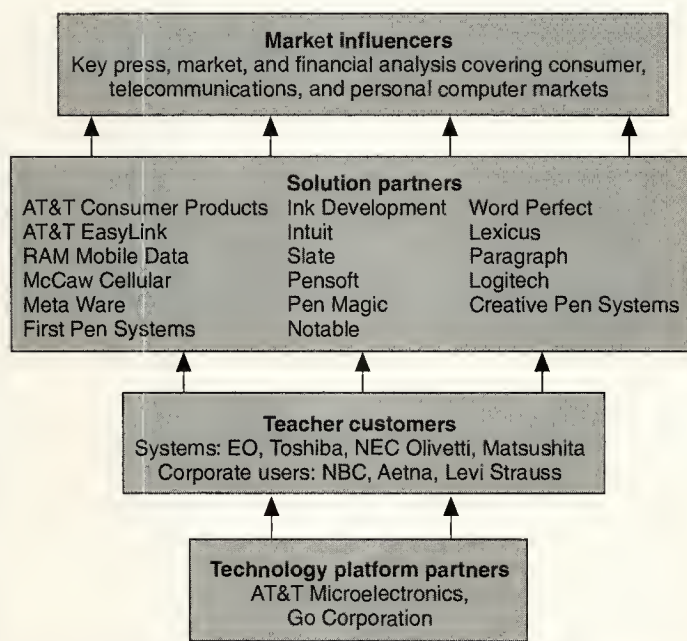

Figure 3. Personal communicator camp (as of early 1993).

> *The successful camp cooperates to match the needs of a well-defined early target user to the price and functional characteristics of the device.*

forum for this was a series of intercompany meetings that included management and key functional team members. The most important decision the group made was defining the personal communicator market as the market that the two companies would jointly pursue and develop. They agreed on a customer needs definition, a price point, and target applications. These decisions formed the strategic basis of the relationship and was the beginning of a market development effort. The Hobbit/Penpoint camp continues to hold strategic meetings and now includes the third key member, EO, the first personal communicator systems manufacturer.

Teacher customer programs

Another important result of those early meetings was that AT&T and Go identified early-adopter target users and defined a recruitment and support program. The Hobbit/Penpoint team reasoned that the early personal communicator system OEMs probably wouldn't come from the conservative rank and file of the personal computer manufacturers. Rather, they looked to start-ups, like EO, and to consumer manufacturers in the world's biggest consumer market, Japan.

They identified 10 early OEM targets, mainly in Tokyo, and defined the recruitment program. Recruitment consisted of a concentrated, top-down conceptual selling effort led by the executives of AT&T Microelectronics and Go. They saw the opportunity to develop the new personal communicator market and change the one-size-fits-all paradigm of personal computing. Their vision depended on the corporate capabilities of AT&T as a market development partner.

Educational seminars at the operational level that dealt with design and development issues followed the executive meetings. Successful early meetings resulted in secured wins at Toshiba, NEC, and others. The team implemented a process to provide strong support to these customers, and importantly, to bring market and technical learning from these customers back into AT&T and Go.

Application solution partner programs

Early customer learning provides the basis for identifying the end-user solution set elements. In the case of AT&T, a different class of use paradigms and applications for a new group of mobile personal device buyers, and a new set of communications technologies, will define the personal communicator market.

The example also shows that market development goals determine how to evaluate and target solution partners. AT&T recruited new, innovative application software partners such as Ink Development to commit to the Hobbit/Penpoint platform. For handwriting recognition and other functions, AT&T recruited algorithm suppliers and development software vendors. For large companies putting solution camps together, some of the most effective relationships are those formed with sister groups and divisions. The Hobbit group within AT&T Microelectronics focused on teamwork with

- other AT&T chip groups for DSPs and modems for communications functions,
- the EasyLink e-mail division for communications applications and services,
- consumer and communications systems groups, and
- the Phone Center group for personal communicator distribution expertise.

AT&T also focused on corporate relationships by integrating RAM Mobile Data and McCaw Cellular into the camp.

Market influencer programs

When the personal communicator market was first begun through the AT&T and Go programs, no experts existed. Instead, the team brought together recognized visionaries and influencers in the areas of personal computing, telecommunications, and consumer electronics. One of the early, most effective programs resulted from recruiting one very influential communications market analyst, one well-known PC analyst, and one financial community guru into the camp. AT&T and Go management engaged the influencers in one-on-one discussions to get input on the name of the new market segment, the technology and market trends that would drive it, and the success factors. The three influencers, in their publishing and speaking activities, acted as lightning rods around which other financial analysts, market analysts, and various consultants collected to study and forecast the personal communicator market.

The success of camp development depends primarily on two factors:

- the camp's ability to assemble a complete solution so that mainstream buyers feel comfortable in new technology adoption, and
- the commitment among all camp members to market development.

Table 1. Competing mobile computing camps.*				
Product	Apple	AT&T	General Magic	Compaq/Microsoft
Operating system	Newton OS	Penpoint	Magic Cap	"WinPad"
Microprocessor	ARM (Advanced RISC Machines)	Hobbit	Multiple: MIPS 3000, Power PC	Intel x86
Communications components	Cirrus Logic	AT&T Microelectronics	Motorola	Unknown
System manufacturers	Apple, Sharp, Matsushita, Motorola, Siemens/Rolm	AT&T, EO Matsushita, NEC, Olivetti, Toshiba	Motorola, Matsushita, AT&T	Compaq, Matsushita
Communications software	Transport solutions in core of OS	Transport solutions in core of Penpoint	Telescript protocol, language of "WinPad"	Transport solution in core of "WinPad"
Services	Apple Online Services	AT&T EasyLink Mail Services	Mead Data News Corp. Ltd.	Microsoft news, information services

*Source: Technologic Research, New York, Apr. 26, 1993

As companies jockey to gain control of a standard and speed the process, they often assemble multiple technology partners but fail to adequately involve such market partners as software, services, and support players who must use the standard and provide end-user value. When this happens, market adoption is retarded and multiple competing camps, each serving a slightly different market segment, begin to form. The results of this scenario are unfortunate for the end users because they find multiple standards just as frightening as a lack of standards.

Unfortunately, this multiple-standards phenomenon already plagues the personal communicator market. An Apple Computer consumer product vision is competing with a General Magic/Motorola information services device camp and a Compaq/Microsoft personal computer extension system (refer to Table 1). As you read this article, AT&T will have integrated Go and EO into its operations, the Apple Newton will have been on the market for months, and others will have shipped similar devices. Who will win? How would we apply camp development concepts to determine which camp might be more successful?

Two principles from this article help in answering these questions. First, the winner will be the camp that can consistently cooperate to match the needs of a well-defined early target user to the price and functional characteristics of the device. So far this goal has been elusive. An electronic product on the near side of $1,000, fully equipped, is a chal-

lenging general consumer product sell, and begs to be compared with the personal computer. Are these devices really targeted at the mobile PC user? The home PC user? If so, why does anyone need one, as opposed to a PC with a pen interface? On the other hand, an under-$500, tightly focused, almost-dedicated device could potentially appeal to the larger market of non-PC users. The obvious risk here is that if the vendor picks the wrong dedicated function, the total market may not yield enough volume. The winning camp will wrestle with this issue and come up with a dedicated, or at least modular, product definition.

Secondly, once the target market is better defined, missing solution pieces must be supplied. Although each camp includes the partners to potentially address most of the missing elements, no camp is delivering

- integration of telephone interface and processing;
- widespread end-user application education and support; or
- testimonials and consistent consumer-quality promotion, packaging, and distribution.

The challenge to AT&T is great. It must effectively harness the efforts of over 35 personal communicator camp relationships, established during the first year toward common market development activities and milestones during the second, third, and subsequent years of camp development.

PARTICIPANTS IN BOTH FORMAL AND INFORMAL standards development efforts can use the checklist in the adjacent box as a guide to assure that the process acts to encourage new technology adoption. These questions assume that the fundamental purpose of standards definition is to achieve new market growth by not only enabling competition but by also addressing the needs and risk objections of new technology users. ▥

New technology adoption checklist for standards participants

How does the process perform against the ultimate quality measures: Are we slowing down market acceptance?
- Has the market already voted on the standard, and, if so, is the process supporting the user's choice?

Are early users represented and effectively involved in the process:
- Innovators, to help design in technical quality?
- Early adopters, to define the application solution set?

Which costs of new technology adoption does the new standard specifically address?
- Which cost components are not addressed?
- Are related and complementary standards necessary?

Are the process and communications about process progress structured to lower the users' perceived risk of adoption?

Is the process trying to deal with more than one application segment?
- Honest, prolonged, and strong disagreement on an interface specification may indicate the need to initiate a separate effort.

Is there a market development plan in place among standards that leverages the standards definition process?

References
1. Chris Halliwell and David Schneer, *Standards Case Study Research Results* (for Pacific Bell), Regis McKenna, Inc. project, 1985; contact author for information.
2. Aaron Goldberg, "EISA: Questioning the Vitality of the Platform," *Computer Reseller News*, Nov. 19, 1990, p. 57.

Suggested readings
LaPlante, Alice, "MCA/EISA Debate Gone; No Forwarding Address," *InfoWorld*, Aug. 27, 1990, p. 20.
Mangione, Paul A., "Who's Flying ACE?," *Unix Review*, Vol. 10, No. 9, Sept., 1992, pp. 29-32.
McKenna, Regis, *Relationship Marketing*, Addison-Wesley, New York, 1991.
Moore, Geoffrey A., *Crossing the Chasm*, Harper Business, New York, 1991.
Morris, Charles R., and Ferguson, Charles H., "How Architecture Wins Technology Wars," *Harvard Business Review*, Mar.-Apr., 1993, pp. 86-96.
Rogers, Everett H., *Diffusion of Innovations*, 3rd ed., Free Press, New York, 1983.

Chris Halliwell is an independent consultant providing strategic technical marketing services to leading suppliers of electronic components and related services. She teaches the Strategic Marketing of Technology Products course in the California Institute of Technology executive education program. Previously, she was a managing partner with Regis McKenna, Inc. She has also worked for Intel Corporation, Rolm Telecommunications, and IBM, and started her career on a National Science Foundation grant in information services.

Halliwell holds a bachelors, a masters in information services, and an MBA, all from the University of California at Los Angeles.

Direct any questions concerning this article to the author at 637 Hawthorn Avenue, Los Altos, CA 94024; or halliw@ netcom.com.

Reader Interest Survey
Indicate your interest in this article by circling the appropriate number on the Reader Service Card.

Low 153   Medium 154   High 155

War of the Words*

Intellectual Property Laws and Standardization

(* With apologies to H.G. Wells' *War of the Worlds*)

While strong economic forces drive the computer industry to develop common standards, the concepts of standards and compatibility inherently conflict with intellectual property laws. Such laws create temporary monopolies as incentives, yet exclusive ownership of standards is inconsistent with the very idea of a common standard. Absent concerted industry action, confusing and often technically illiterate court decisions relating to these issues may ultimately encourage litigation, rather than standardization efforts.

G. Gervaise Davis III

Davis & Schroeder, P.C.

If you have ever tried to connect one electronic mail system to another with high-speed "Hayes-compatible" modems that aren't compatible, using RS-232C "standard" cables that won't interconnect, you know how intellectual property lawyers feel these days. Computer industry clients often ask for our help in avoiding litigation as they develop software intended to be "compatible" with some "de facto" industry-standard software package or collection of "standard" hardware. In today's litigious world, that's not an easy task for the lawyer or the client.

How did we get here? This article discusses how the industry got itself into this legal fix and where it is going, based on recent US court cases. That may not sound like a must read, but this situation should concern you nonetheless. One of these days, it may get you or your employer sued for copyright infringement, for doing things any rational computer scientist or engineer would consider the normal way to write computer programs or build hardware. While the cases discussed largely involve software, please understand that these legal decisions may have considerable effect on hardware designs. Hardware today contains lots of embedded software. If you find the conflicting court decisions and inconsistent legal rules confusing, take heart—you're not alone. The courts and lawyers are confused as well.

Everyone in the computer industry understands that all technology is based on incremental progress. Using the ideas of others to improve old products or create new ones is a time-honored tradition. We understand further that computerized devices can talk to each other only if a common interface or set of rules lets them. After all, one person speaking French and another English can hardly expect to communicate successfully without a translator. Similarly, users of various software packages have come to expect certain standard features, such as the use of the F1 key for Help or the ESC key to back out of a screen. Sound economic and business reasons compel everyone to follow these practices, and most industries are built on this approach to product improvement. However, this practice requires that copyright law recognize the need for common standards and for compatibility between systems and software. Unfortunately, the law presently does not do so, for reasons I will soon make clear.

The bad guys. Sadly, a small group of lawyers and their well-funded, well-established clients think that the intellectual property laws should work to prevent innovators from making compatible software and hardware. Not wanting standards, they fight to maintain rigidly proprietary systems and equipment, an approach that locks out competition. They do not want products that work similarly, especially if their product has

> *Companies that oppose standards often seek to protect their bottom line with phalanxes of lawyers. They sometimes find it cheaper to litigate than to improve their products.*

become the de facto standard that others must meet. Reverse engineering and compatibility are forms of theft in their eyes, and they care about neither standards nor the economic effect this approach has on their customers.

The legal issue facing the industry, then, is whether software-based hardware devices, or software itself, should be given very broad protection under the copyright laws. Or should the scope of such protection be narrow, to encourage improvements and new developments that build upon earlier work? There is little agreement on the answer.

For example, Anthony L. Clapes, an IBM patent attorney, recently took the courts and lawyers to task for their "errors" in trying to limit the scope of copyright by allowing reverse engineering and permitting compatible software.[1] In *Softwars: The Legal Battle for Control of the Global Software Industry*, he contends that permitting any form of compatibility is "weakening the copyright laws," and a serious error on the part of the courts. *[See the MicroLaw column, pp. 78-80, in the June 1993 issue of* Micro *for a review of Clapes' book.—Ed.]* Companies that oppose standards often seek to protect their bottom line with phalanxes of lawyers. In my judgment, they sometimes find it cheaper to litigate than to improve their products. They rely on these laws, sue people, and cause smaller defendants to spend a ton of money defending what you and I would think were impossible cases. I know. I have defended some of those crazy assertions, and sometimes we have lost.

Courts ill-suited for complex scientific issues

The federal courts and the US adversary legal system generally, which we inherited from England, are poor places to decide whether one scientist misused another scientist's software, designs, or ideas. A trial simply does not allow sufficient time to explain the technical issues adequately. Trials are adversary proceedings where each side argues its own

best interests. Often the interests of the industry as a whole are never even considered.

In such cases, unless someone else in the industry files an *amicus*, or friend of court, brief, the trial judge is completely unaware of the broader consequences of the dispute. Also, the cost of copyright or patent litigation is terribly high by any standard. A typical copyright case can cost each side $500,000; patent cases can go on for years and cost millions. Worse yet, the court caseloads are growing every year, while funding for the courts gets cut back like everything else these days. Courts are just not the ideal place to determine these complex policy decisions.

In a war of words—a lawsuit is fought with legal and technical words as weapons—even the best intentioned federal judge has trouble understanding the terminology. What computer scientists mean when they say someone stole their GUI, swiped their interface specifications, or purloined their algorithm frequently gets lost in translation. Technical experts, paid to take a particular positions, often don't help. They insist on explaining everything in computerese and dwell on their own hyper-technical areas of expertise. To compound the problem, lawyers and judges also use their own arcane language as a convenient shorthand, just like computer folks do. In both cases, as the song says, "the words just get in the way."

A lot of other things get in the way, too. Sometimes a lawyer's job involves cross-interpreting this confusing babble from one group to the other. The lawyers themselves may have a poor or nonexistent understanding of the technical issues, either because they did not take the time to learn the information or because their clients did a poor job of explaining the problem. Considerable time in computer industry cases is spent educating the judge and jury on the technical issues involved. We lawyers need the help of bright people in this industry to cut through all this background noise, or the clutter will obscure the main signal.

Ideas belong to everyone

Understand, also, that there is an inherent conflict between the philosophical and scientific concept that ideas should be freely used by all, versus the whole structure and purpose of the intellectual property laws. Those laws are a form of temporary monopoly governments grant for reasons of public policy. Thomas Jefferson made the powerful analogy that other people using his ideas harmed him no more than if someone were to light their candle from his. He, most of all among America's founding fathers, fought for the concept that ideas belong to everyone. Yet Jefferson, nonetheless, saw to it that patent and copyright laws came into existence. He and like-minded architects of the Constitution agreed on a specific provision authorizing patent and copyright laws, a provision that is still the basis for all these laws and cases. Since Jefferson's day, the Congress has passed numerous

intellectual property laws, starting as far back as 1790 with the first copyright law that protected maps, charts, and books from copying. The early patent laws also stem from this constitutional basis.

As US Supreme Court cases have explained for years, these laws create a limited monopoly for the author or inventor, in exchange for disclosure of the information in the book or concerning the invention described in the patent. Theoretically, people would not disclose this information if they could not protect it and make money off it, a lack of disclosure that would seriously deprive science and the arts. The primary purpose of copyright laws, the courts have said, is to encourage this disclosure, not to protect private property or the income of the inventor.

For this reason patents expire after 17 years, a period that made sense when scientific knowledge doubled every 100 years, instead of today's doubling every three years, if not every three months. Copyrights, conversely, last for the life of the author plus 50 years, or 75 years if owned by a business whose employees prepared the computer program. Considering the degree of obsolescence of most software and hardware, this is also too long.

Copyright law is a poor way to protect software

Just as this 18th century process to encourage disclosure has become somewhat obsolescent with the ever-shortening time value of scientific knowledge, the US likewise made a fateful decision in 1981. Congress decided, for geopolitical reasons, to protect computer software under the copyright laws, rather than to create a specific new body of law that recognizes software as a vastly different form of written work. Congress wrongly analogized software to books and other written works. Because it is a writing rather than a work of art, photograph, or sound recording, computer software is protected under copyright as a "literary work." Though we know now that this is wrong for lots of reasons, it is probably too late to correct the situation. We will have to work within the system to fix it. Congress used copyright law to protect software because we already had treaties with many countries that protected our copyrighted works. This approach was easier than negotiating new rules for a new kind of legal rights in software.

Copyright does not protect function. As author John Hersey wrote in his dissent to the 1979 CONTU report to Congress, this decision to protect software with copyright laws was a mistake because historically copyright has never protected ideas (see the adjacent box for a description of CONTU and its activities). Nor has it protected a "procedure, process, system, method of operation, concept, principle, or discovery, regardless of the form in which it is described, explained, illustrated, or embodied" in a copyrighted work. In fact, Title 17 USC, 102(b) of the copyright law states exactly this same prohibition. How then can you protect a com-

puter program, which is inherently functional because it "does work" as Hersey argued? Until such time as the Supreme Court rules on the scope of copyright protection, the answer will be with great difficulty and a lot of expensive and pointless litigation. The whole problem stems from the very character of software, which is a written description of a process or system intended to be used by a computer. Software is, in effect, an instantly reconfigurable machine, made of human words, that the computer recognizes when converted to binary digits for it to "read."

Today, no one seriously argues that the actual source or object code of a computer program can be copied with impunity. It cannot be, and to do so is infringement. This is now well-established law. The key issue now is what beyond the literal code is protected in a work such as software, which is inherently a description of a method for doing something? The real question is whether two programs that function the same way, but that use entirely different code (and that was not merely paraphrased one from the other), can be developed and marketed without one being considered an infringement of the first?

The issues being litigated today are:

- How much graphical similarity can there be in the screen interface to the user (*Apple Computer* v. *Microsoft/HP*,[2] now in the appeals process)?
- How much functional similarity can there be between two programs (*Autoskill* v. *National Educational Support Systems, Inc.*,[3] now on appeal to the Supreme Court)?
- What can you legally do to obtain information needed to write programs that will actually talk to another computer without getting into trouble (*Sega Ltd.* v. *Accolade, Inc.*[4] and *Computer Associates* v. *Altai*,[5] from the Ninth and Second Circuit Courts of Appeal)?
- Are data structures protected by copyright (*Computer-max* v. *UCR*[6])?

Gates Rubber

In a case decided in October 1993, the Tenth Circuit in *The Gates Rubber Company* v. *Bando Chemical Industries, Limited*, Case No. 92-1256, reversed a lower court determination that a mathematical constant involved in a formula for sizing fan belts was copyrightable.

In that decision, the court discussed in some detail the fact that most processes are not copyrightable, even though they might be implemented in a computer program. This case represents a new hope that the courts will begin to understand the problem of processes and standards.

- Is there any right to compatibility, and how far does that right extend (*Nintendo* v. *Atari*[7])? The Gates Rubber box describes a brand new case that has bearing in this area.

Standards are based on similarity; similarity is infringement. We finally get to standards, you say! Yes, and the problem all such cases squarely face is that unless a standard is so well accepted that no one can seriously claim ownership under copyright, we still have difficulty deciding when something is such a de facto standard that it can be used. We may be years away from learning the answer to this question, and there are precious few guidelines at this point. Recent decisions like that of the trial court in *Lotus* v. *Borland*,[8] only make matters worse.

Where a plaintiff cannot prove actual, "direct" copying of the source or object code, he or she can only win the case by showing that the defendant had indirect access to the program (saw it running or read its source or its documentation) and that the second program is "substantially similar." Substantial similarity is, therefore, an inferential substitute for the lack of proof of direct copying. Trouble arises because programs that rely on the need to use standard interfaces or standard functions, by definition, will always be substantially similar. The purpose of a standard is to make it easier to use by following the same rules the earlier designers did. This is the proverbial Catch 22. Do you begin to see the inherent conflict between copyright and standards?

The *Lotus* cases. As noted earlier, no one could seriously claim today that any computer program protects the use of F1 for a Help key or the ESC key to back out of a program. There may be no rule that says you have to do things this way, but nearly every computer user will expect a program to work that way. Yet Lotus, in its several cases against clones and work-a-likes, has convinced the trial judge that unauthorized use of their command trees (which include such common, one-word commands as Print, Quit, and File) constitutes infringement. In the most recent *Lotus* decision by Judge Keaton, Lotus has convinced the court that any form of conversion by Borland of 1-2-3 macros and commands on the fly is an infringement, since it involves Lotus's use in another program. The decision does not even make it clear that the court would approve a one-time conversion at startup. This decision is extremely dangerous to any efforts at standardization, and the industry should watch this appeal very carefully. With any luck, industry members will consider filing *amicus* briefs to educate the First Circuit, which will hear the appeal.

Lotus has also successfully argued that use of an @ to start a formula is something they own, since their Lotus 1-2-3 computer program used it first. The trial judge appears to have concluded that even the macros they use, and the way they are used, are protected. At least the trial judge conceded that the spreadsheet "L" shape was not protected, nor was the concept of numbering or lettering the rows and columns. Conversely, the trial judge in the Apple/Microsoft dispute, which involved the similarities between the Macintosh operating system and MS Windows, rejected many similar contentions by Apple. Apple, in that case, claimed ownership of overlapping windows and the use of icons in place of word commands, and made many other rather startling assertions to anyone familiar with any windowing environment. Fortunately, Judge Walker severely limited Apple's claims. Apple has now conceded defeat in the lower court, but is in the process of appealing.

Using 1,2,3 to number answers may be infringement. By contrast with the Apple decision, in the recent *Autoskill* v. *NESS* case, an appeals court has recently ruled that numbering answers "1, 2, 3" and requiring the user to press those same number keys was one of the elements that made the second teaching program an infringing work—a judgment that put the defendant out of business. That same court held that the second program to teach reading could not be organized so that it first tested the students, then analyzed the results, then taught them, and then tested them again on a timed basis. That process was substantially similar to the organization of the first program. As one of the witnesses said, "The two programs do the same thing, they function the same." Both the Trial Court and the Appeals Court have now held that to be infringement. The Supreme Court, which does not have to hear the appeal, recently declined to do so.

Autoskill demonstrates that even when applying common-sense standards, such as when there is no other practical or normal way of doing something, the present crazy-quilt of the law may be dangerous to you and your software and hardware designs. You may have to prove that even though your program works similarly to another program, you did so by means other than copying it. This can be a difficult

thing to show a nontechnically educated judge. Incidentally, both the reading programs in this case were based on a series of educational research papers that taught educators which educational techniques worked in teaching children with reading difficulties and which did not. The papers described a reading disability test and remediation method, which the plaintiff Autoskill then implemented in a program. The *Autoskill* decision would seem to say that if a particular process works for someone else, you can't use it unless you are the first one. Apply that to the design of your next program and see where it leads you!

Protecting data structures. Recent cases have presented even worse examples where the courts have totally misunderstood the programming reasons for similarity, and found the second program to be infringing. In *Computermax* v. *UCR*, a recent Georgia case, the trial judge found infringement primarily because the data structures of the two client applications programs were the same for business data intended to run on the same host/server machine. The court failed to realize that the data—common business information such as names, addresses, and part numbers—had to be read out and into the host system, even though the client systems used different software. Plaintiff's counsel managed to convince the court that in such cases different data structures should be required and that the data had to be converted to be used. It is as though once a particular data structure is used, no one else can use that structure. Instead, they must create their own, no matter how inefficient that might be from a programming and resource standpoint. Unfortunately, this case was settled without appeal, because the cost of an appeal was uneconomic for the defendant.

Light at the end of the tunnel

Recent court decisions offer occasional glimmers of hope. In *Computer Associates Int'l* v. *Altai, Inc.,* the appeals court approved a lower court decision that held Altai had not infringed upon a copyright by developing a computer applications program which functioned similarly to an earlier one owned by Computer Associates. Both competing programs worked in the same IBM mainframe environment and with IBM operating system software. Both companies developed design specifications that considered the operating environment and required nearly identical programming interfaces.

The *Altai* court recognized, for the first time, that the Altai programmers were constrained in the design of their software by standards and protocols established by IBM, which forced the second applications program to have a number of similarities to that of Computer Associates. Thus, if you can show you must interface with another program or other hardware, you may not be precluded from programming those elements that are similar because of the need to interface with a fixed standard. You cannot use the same code, however, and must write your own. In other words, if you can

---

In none of the cases was there a charge that the actual source code was similar, only that either the organization or the screens presenting an interface to the user were the same.

---

show that it is functionally necessary to do things the same way, you should be all right, if you write your own code to perform those functions.

Whelan and *Plains Cotton* distinctly differs from an earlier, much criticized, case. In *Whelan Associates* v. *Jaslow,*[9] the courts held that similar "structure, sequence, and organization" of two programs was almost always going to lead to a finding of infringement, as it did there. That decision, in the judgment of many courts and lawyers, failed to consider the inherent similarity in all accounting programs intended for the same industry, and the fact that the same person designed (but did not write) both programs. The decision also displayed a severe misunderstanding of how the organization of a program—its various modules—may be dictated by common practices and common problems.

About the same time the Whelan case was decided, another appeals court held, in *Plains Cotton,*[10] that it was not an infringement for the screens and the functions of two cotton trading programs to be very similar. The court said that the terminology of the industry and its practices (akin to computer industry standards?) were so well established that two programs to accomplish the same purpose were not infringingly similar. Note that in none of the cases I discuss here was there a serious charge that the actual source code was similar, only that either the organization or the screens presenting an interface to the user were the same. These decisions are all illustrations of the type of arguments that usually arise when contrasting the protection of so-called "literal" similarity versus functional or "nonliteral" similarity. The cases are not consistent.

Video games and common interfaces

Another compatibility issue that has produced major litigation in the past few years relates to the right to reverse engineer a program to determine how it works, so you can write a program to use the same hardware platform or to "talk" to the first software. Aside from the now established

right to reverse engineer "when there is no other [practical] way to get the information" needed to work with existing hardware or software from other manufacturers, unresolved issues relating to compatibility remain. The two key questions are:

- How far can you go toward making things like your video game cartridge function with the video game hardware of other makers?
- What, if anything, can you do to anticipate things that could later be done to lock you out?

Two recent California cases illustrate these problems rather graphically.

Lock-out devices. Nearly all video game consoles for home use employ electronic security devices intended to lock out game cartridges made by other than "authorized licensees," who must pay stiff royalties for the right to develop and sell compatible games on these platforms. These security devices are either software or a combination of software and hardware in the console (or base unit). They are not documented anywhere, for obvious reasons. All require that the game cartridge "talk" to the base in a predetermined manner. If it does not, the game base shuts down and ends the game.

Sega v. Accolade. In this case, Sega stored a program in the game console that jumped to a particular location in the game cartridge ROM and required the data, "SEGA," to be in that location before the game would continue. Without this data, the console or base would not run the cartridge. Accolade had made games for the Sega base for some time, until its games suddenly stopped functioning, when used with the newest Sega bases. Since Sega would not explain to anyone what it had done, Accolade reverse-engineered a number of Sega bases and cartridges to see how, if at all, their game code differed from the Accolade setup code. Accolade discovered the nature of the security device and put the SEGA data in the proper location and also developed setup code that was similar, if not nearly identical to Sega's. Sega sued Accolade for copyright, trade secret, and trademark infringement, and succeeded in obtaining a preliminary injunction in the trial court.

On appeal, Accolade, with the help of several software industry *amicus* briefs, explained to the court that the Sega technique was a form of an interface standard, which had to be used or the competitive cartridges would not work. Accolade argued that when a copyrighted work is used to create a standard or protocol that cannot be worked around, the subsequent software developers working with this platform must have a right to use the minimal amount of data or code necessary to make their product work. In this sense, it is an extension of the conclusions of the court in *Altai* that allowed similar interface code because of a common plat-

form. A major difference here, however, is that the Accolade product was intended to replace the Sega cartridge, while still using a Sega base from which Sega could still derive revenue. By contrast, Altai was selling a program that replaced the Computer Associates program, but both programs ran on IBM hardware and under software standards set by IBM. In other words, Sega set itself up for the problem. Computer Associates did not.

Reverse engineering. The appellate court in Sega also gave limited approval to the concept of reverse engineering, which is vital to anyone interested in standards. There is often no other practical way of obtaining the required interface information without reverse engineering. IBM, Apple, and DEC, among others, have contended for years that reverse engineering of another's program is copyright infringement because a copy of the original program must be made to read and understand it, so that the necessary interface information can be extracted from it. The Ninth Circuit rejected this argument, ruling that where the interface is not published and there is no other reasonable way to find the information needed for compatibility, reverse engineering is permissible. The court cautioned, however, that the final (second) program must not be a copy of the first. It must be independently written to meet the functional needs of compatibility, but must otherwise be dissimilar except where required by the standard being followed.

Nintendo v. *Atari*—The limits of compatibility

About the same time that Sega was in the trial courts, the same trial judge had originally ruled in *Nintendo* v. *Atari* that reverse engineering was illegal and was an infringement of the copyright laws, and issued a preliminary injunction against Atari. She was later reversed on this issue by a higher court and the case went back for reconsideration and actual trial. Here the facts were similar to Sega, but the lockout device was far more subtle and intractable.

Ultimately, the trial judge approved reverse engineering as a practice, but only where necessary to understand an interface specification. The judge, however, still found Atari guilty of copyright infringement because it had not only implemented the same method of responding to the lock in the Nintendo cartridge, but had also attempted to anticipate other possible means that might be used to lock its products out. To do so, Atari coded in an unused function in the Nintendo base code. Understand, however, that Atari did not use the Nintendo code at all, but wrote its own functionally similar microcode for a completely different microcomputer used in its own cartridge.

The Nintendo video game system, in its initial version, had an identical lockout chip in the base and in its cartridge. Both had an elaborate handshaking scheme by which the base sent a precisely calculated signal to the game cartridge. The cartridge, in turn, made some mathematical calculations and

returned a very precisely timed stream of bits to the base. If the base did not receive exactly the signals it expected, it went into a loop and did nothing. While the cartridge had the same program to check the accuracy of signals it got from the base, it had no reason to shut the base down, so it did not test for the accuracy of the signals from the base. In later game cartridges, Nintendo eliminated this superfluous aspect of the program stored in ROM.

Atari reverse-engineered the base and the cartridge to view the original programs (and the later cartridge programs). It concluded that if its own cartridges could not check the accuracy of the signals from the base, simple changes in the Nintendo base program could later lock Atari's cartridges out. Atari therefore included the function that checked the base signals, but did not permit the program to abort the base if the signals did not match.

Providing for future compatibility is illegal. Because of this precaution, the trial judge granted judgment to Nintendo for copyright infringement by Atari. While Atari had the right under copyright law to include all the absolutely necessary functions to make its cartridges work with the Nintendo base, it had no right to anticipate future designs of the Nintendo base-cartridge interface. Inclusion of these future compatibility features was "evidence of copying" of the Nintendo code, according to the ruling, even though the code was not the same—only the functions were.

This case is still pending in the courts on various patent issues and undoubtedly will be appealed, if not settled, somewhere along the line. However, it is disturbing that the trial court effectively prevents anyone from ever doing anything to anticipate changes in an interface. It fails to consider that interfaces are changed from time to time, or that Nintendo, itself, had changed its own cartridge program. It also raises the interesting issue that if a plaintiff wanted to strengthen its own case, just before filing suit for infringement, it could subtly change its program by deleting some heretofore unused function. It could then complain that the defendant's code included "unnecessary functions." This ruling requires that a potential defendant in a work-alike software case constantly reverse-engineer the possible plaintiff's product to make sure that no changes have been made that would give rise to such a charge. It simply does not make sense. Subsequent proceedings will clarify this aspect of the *Nintendo/Atari* case.

What does this all mean to compatibility and standards?

Present case law interpreting the scope of copyright protection is still in a state of uncertainty that likely will not be resolved until the Supreme Court takes one of these lower court cases and provides its own, final interpretation. Under US law, the various Circuit Courts of Appeal, of which there are 11, create the law for the states within their jurisdictions.

> *Copyright is not supposed to protect function, but the courts clearly disagree as to what constitutes unprotected function.*

Conflicts have arisen between several of these circuits, and there are many inconsistent cases in the trial courts, among both decided and pending cases. For the present, only a few fundamental principles can be divined in this area, and even these are murky. A recent study by the Congressional agency established to look at technology issues, the Office of Technology Assessment,[11] discusses these murky issues. It concludes that Congress or the courts will have to make a policy decision on how far the scope of copyright is supposed to extend. For now, it is anyone's guess.

The risk of litigation exists. Programmers and their employers developing functionally similar programs need to have a well-developed rationale for what they do when developing compatible programs. Otherwise they risk copyright litigation with the owner of the earlier program. Copyright is not supposed to protect function, but the courts clearly disagree as to what constitutes unprotected function. The problem is that noted at the outset—lack of time in a short trial, lack of understanding of the technical issues by the courts and some of the attorneys handling these cases, and most of all, lack of any uniform position by the computer industry on what the scope of protection should be.

In theory, two computer programs that do exactly the same thing, exactly the same way could be written and marketed by two independent companies without involving infringement—as long as neither had access to the other's program or detailed documentation before the release of the second of the two. Once the element of access to the source code of the first program is established, the rules get more confusing. Clearly, reading the source code of the first program would provide a detailed understanding of how and why things work as they do in the program. If this were used to develop a structurally, entirely different program, there would be no problem.

But what if the same person works for one employer and designs and writes a program, and then that employee joins a second company where he or she is asked to develop and write a second, functionally similar program? This situation raises thorny problems not easily resolved. Aside from issues

> ### *Engineers may find themselves unwanted if they know too much about their former employer's programs.*

of failure to keep trade secrets inviolate, there may well be a presumption of a court that any similar program written by the programmer is based on the same code, no matter how different it actually is. This was to some degree the fact that confused the *Whelan/Jaslow* court into prohibiting programs with similar structure, sequence, and organization. It bears careful consideration if you or your company get into this situation. Many pending court cases involve this type of complex fact situations. The hazard is that engineers may find themselves unwanted if they know too much about their former employer's programs, since such knowledge can be dangerous to the second employer.

What the law actually says. If Section 102(b) of the copyright law means what it says, even if the second program's author saw the user documentation and saw the first program running on a computer, the programmer should be able to write his or her own program that functions the same, without this being infringement. Functionally similar programs will always look similar in some respects. The legitimate risk should be that the screen messages and graphical elements should not be the same, unless justified by established standards or industry practices or terminology, as in *Plains Cotton*.

Some suggestions. Here are some suggestions that may help you understand what type of similarities can exist with fewer problems under the copyright laws. First, where you can show, as in the *Nintendo* and *Sega* cases, that something will not work unless it is done the same way, you can rely on what lawyers call the "idea/expression dichotomy" or the "merger doctrine." Ideas are not protected by copyright, but the way they are expressed in writing is. Therefore, if you can show that from a programming standpoint there really is no other practical way to express a formula, for example, you can rely on a "merger" defense. The idea has merged into the expression, so you can use the same expression. Beware of the judge who says, "Well, weren't there at least one or two other ways to do it?" This is not uncommon, but it should not be the test.

Second, in some cases if you can show that what you did in your program is simply common programming practice, taught in school and documented in textbooks, you should

be able to rely on what the courts call *scènes à faire*. This term is legal French for something like common plots in plays, which everyone uses. For example, "Boy meets girl, they fall in love, and then the parents or their culture make them miserable." Every author will use that plot once in a while. Similarly, binary sorts in database programming are established practices, found in every textbook on the subject. The problem, however, that may arise is that a common programming practice is used, but in an unusual place, or that your common practices are modularized in the same order as the other program. Be prepared to explain why you used a particular technique, if the first programmer also used it.

Finally, there is a concept in copyright of originality. In copyright law this is not the same as the lay meaning. Essentially it means that if you can show that an idea has been implemented by others in many similar ways, you may be able to claim that the screen elements or features used in the first program are not original, and thus are unprotectible. The concept is akin to that of *scènes à faire*, but slightly different. Microsoft successfully argued this in the Apple/Microsoft dispute, especially with regard to Apple's claims relating to the use of windows, overlapping windows, and related elements found in both MS Windows and the Macintosh operating system. It was also applied to defeat Apple's claim to the use of inverse video to indicate selection. Be careful, however, since a unique combination of nonoriginal elements has been held in software and database copyright cases to constitute an original work that is protected by the law.

As these are very subtle defenses in the copyright field, especially in the case of computer software, you need to consult an attorney with substantial experience in this field to be certain that you are home free in a given situation. What makes sense in the programming industry, unfortunately, does not always translate into noninfringement in the Alice in Wonderland world of intellectual property law.

THE MOST IMPORTANT CASES TO WATCH FOR IN THE next two years are any appeal from the Apple/Microsoft litigation, the final decision in the Lotus/Borland case, and the pending request from NESS that the US Supreme Court hear an appeal from the lower appeals court decision. Any of these cases could result in a decision from an appeals court that will bring more order to what is still uncertainty. Such decisions are likely to take a year or more. In the meantime, you should also watch what is going on in Congress on these issues. There is a great deal of lobbying and possible legislation being discussed that would outlaw reverse engineer-

ing and attempt to severely limit the industry's ability to develop compatible software to meet a common standard or interface. The big players would like to lock out new competitors; they will use any legal means available, including changing the laws through lobbying.

This same legislative effort was fought hard in Europe over the past five years and resulted in Common Market legislation that legalized reverse engineering, if for purposes of "interoperability." Interoperability refers to those situations where software or hardware must communicate with other software or hardware, based upon some standard or protocol. If it is not a published or available standard, reverse engineering may be used in Europe to ascertain the interface characteristics and specifications. Beyond that, it is not clear in Europe, either, as to how much further one may go toward ensuring compatibility with a de facto standard or in preparing programs that are functionally similar to another.

Where does this leave the software industry? This inconsistency in court decisions leaves everyone very uncertain but hopeful, just like the intellectual property attorneys who work for your companies and with you. It is a frustrating situation that none of us would have chosen. We can only wait, hope, and litigate. Perhaps, I will see you in court. ▯

References

1. A.L. Clapes, *Softwars: The Legal Battle for Control of the Global Software Industry*, Quorum Books, Westport, Conn., 1993.
2. Apple Computer, Inc. v. Microsoft Corp., CCH Copy. L.R. 627,115 (D. N.D. CA. 1993).
3. Autoskill, Inc. v. National Educational Support Systems, Inc., 994 F.2d 1476 (10th Cir. 1993).
4. Sega Enterprises, Ltd. v. Accolade, Inc., 977 F.2d 1510 (9th Cir. 1992).
5. Computer Assoc. Int'l, Inc. v. Altai, Inc., 982 F.2d 693 (2nd Cir. 1992).
6. CMAX/Cleveland, Inc. dba Computermax v. UCR, Inc., 804 F.Supp. 337 (D. M.D. Ga. 1992).
7. Nintendo of America, Inc. v. Atari Games, Inc., 975 F.2d 832 (Fed. Cir. 1992) and the subsequent District Court decision, 1993 CCH Copy L.R. 627,114 (N. D. CA. 1993).
8. Lotus Dev. Corp. v. Borland Int'l, Inc., 1993 US Dist. Ct. Lexis 10985 (D. Mass. June 30, 1993), supplemental opinion, 1993 US Dist. Ct. 11311, and earlier decision, 799 F. Supp. 203 (D. Mass. 1992).
9. Whelan Assoc., Inc. v. Jaslow Dental Lab., Inc., 797 F.2d 1222 (3rd Cir. 1986).
10. Plains Cotton Coop. Assoc. v. Goodpasture Computer Service, Inc., 807 F.2d 1256 (1987), cert. denied, 108 S.Ct. 80 (1987).
11. *Office of Technology Assessment (OTA), Congress of the United States, Computer Software: Copyrights, Patents, and the Challenge of Technological Change*, Global Engineering Publications, Irvine, Calif., 1993.

Additional readings

J. Farrell, "Standardization and Intellectual Property," *Jurimetrics J.*, Vol. 30, No. 1, Am. Bar Assn., Chicago, Fall 1989, pp. 35-50.

J. Farrell and G. Saloner, "Installed Base and Compatibility: Innovation, Product Preannouncements, and Predation," *Am. Economic Rev.*, Vol. 76, No. 5, Am. Economic Assn., Nashville, Tenn., Dec. 1986, pp. 940-955.

G. Gervaise Davis III is a business and intellectual property attorney. For the past 36 years he has advised Silicon Valley and multinational computer industry businesses on the subject of developing, protecting, licensing, and acquiring software and software-based hardware. He is a founding principal of the law firm of Davis & Schroeder, Monterey and San Jose, California. He serves as an adviser to a number of national law journals in the area of intellectual property, and is a director of the Computer Law Association, a multinational group of lawyers working in this field.

Direct questions concerning this article to the author at Davis & Schroeder, P.C., PO Box 3080, Monterey, CA 93942; (408) 649-1122; gervaise@delphi.com.

Reader Interest Survey

Indicate your interest in this article by circling the appropriate number on the Reader Service Card.

Low 156   Medium 157   High 158

Standard Setting in the United States: Public and Private Sector Roles

Standards are ubiquitous, affecting our lives in a multitude of ways. Because the economic and social stakes in standards are so large, the standards development process must be fair to prevent any single interest from dictating the outcome. Equally important is the relationship between the public and private sectors.

D. Linda Garcia

US Office of Technology Assessment

Standards affect our lives in many ways. Food and drugs must comply with health standards; cars use standardized, interchangeable parts; work places have safety standards; clothing comes in standard sizes; jobs are evaluated according to performance standards; telephones have standard interfaces; and bed sheets are sized to fit standard mattresses. Even our lives have become standardized through our reliance on technology.

Because the economic and social stakes in standards are so large, how standards are set is a matter of some concern. The standards development process must be fair to prevent any single interest from dictating the outcome. Equally important is the relationship between the public and private sectors.

In the United States, almost half of all standards are set by the private sector as part of a voluntary consensus process in which all the key players—including government—participate. The system reflects American political culture, and the general preference for market-based, pluralist solutions. However, because standards serve both public and private functions, this arrangement has not been without tensions. And every so often these tensions have erupted from under the surface, as is clearly the case today.

The current US standards process was adopted at the turn of the century, as the nation entered the industrial age. Its form reflects American political culture and the manner in which industrialization took place. In contrast to many other countries, where unified national standards bodies were established in conjunction with the state, standards development organizations in the US first emerged in the private sector, in response to specific needs and concerns.

Today, however, the US economy is in a state of flux due to a number of developments. These include the emergence of a highly competitive global economy in which the US is no longer dominant, the rise of regional trading blocs, the growing importance of multinational corporations and other transnational, nongovernmental institutions, and the rapid advance of technology. Just as the industrial era gave rise to the present standards development system, so too these structural changes are placing new demands on it, raising questions about whether a new balance must be struck between the private and public sectors' roles.

Standards evolution and the key players

Economically motivated standards have proliferated and become more highly valued, as economic relationships have became more intricate. Mass production meant standardized processes that required standardized parts. The demand for interoperable parts was especially prominent in the US where the economic conditions for large-

scale production were ripe. In no other country was there a geographic market large enough to absorb the output of a standardized commodity nor a demand stable enough to sustain continual large-scale production. Nowhere was a labor or consumer market as large as that in the US, which could take advantage of an ever-expanding volume of mass-produced capital and consumer goods.[1]

Standards were also spurred on in the US by the extension of trade across the continent. As trade became more dispersed, standards were needed to assure that products manufactured in different locales could work together and be easily replicated, assembled, and repaired. Moreover, standards were required to facilitate trading itself. For example, the railroad extended trade over vast regions, so procedures for billing and exchange were also standardized through bills of lading.[2]

As the importance of standards increased, so did the number of people who had a stake in the selection of standards. Producers got involved in standardization when trade was extended across greater distances. Standards served as a trademark, allowing producers to differentiate their products from their competitors, and to price products for different markets. It was to this end, for example, that American farmers played such an important role in setting agricultural standards during the first half of the 18th century. They realized that by grading and classifying their products, they could set up separate distribution channels and increase their profits. Thus, when farmers moved west, they labeled their products by the region of origin, while wholesalers used these names—Goschen butter, Genessee flour, and Herkimer cheese—as designations of grade.[3]

Suppliers were brought into the standards process with industrialization and the development of precision manufacturing. Recognizing that production costs could be greatly reduced with interchangeable parts, they began to produce to specifications. Gun manufacturing was one of the first industries in the US to take advantage of production based on interoperable parts, followed by clock making and the manufacturing of bicycles and sewing machines. In 1813, Simon North signed a contract with the Federal Government to produce 20,000 pistols. His contract specifically stipulated that "the component parts of pistols, are to correspond so exactly that any limb or part of one pistol ... may be fitted to any other of the twenty thousand."[4]

Consumers also gained from standardization. Mass-produced goods were cheaper. Thus many consumer goods—such as cars, refrigerators, and vacuum cleaners, which were once regarded as luxuries—became more accessible to all. Between 1914 and 1924, Ford produced more than 15,000,000 standardized Model T's, the cost of which dropped during the same period from $950 to $240.[1]

Standards also conveyed product information and provided greater quality control. One of the first product areas to benefit from standards was that of food. Responding to scandals

---

As the importance of standards increased, so did the number of people who had a stake in the selection of standards.

---

in the meat packing industry, Congress passed the Pure Food and Drug Act of 1906. This legislation not only protected against misbranding and food adulteration; it also standardized containers for marketing fruits and vegetables, thereby eliminating false measurements and deceptive shapes. Later the Department of Agriculture, continuing the standards program initiated during the First World War, developed standards for fruits, vegetables, peanuts, honey, butter, cheese, eggs, and meat, and established inspection stations at a number of key distribution centers.[5]

The general public became even more attuned to the need for standards because of the many problems accompanying industrialization. With more and more mishaps due to the rapid expansion of technology, safety standards were introduced. An average of 1,400 boiler explosions per year led the American Society of Mechanical Engineers to write a comprehensive boiler code in 1910. Once most states and cities had moved to adopt the code, such explosions were virtually eliminated.[6]

With the advance of technology and its further deployment in industry, scientists and engineers began to play a special role, as a group, in standards development. Faced more and more with the need to quantify their results, they could not proceed in their work without more accurate standards of measurement and precision instruments to take these measurements. Thus, even though standards were a boon to industry, it was the scientists and not the industrialists who called for national standards to be developed through a Federal Bureau of Standards.[7]

Although the federal government became involved in standards as early as the mid 1980s through the work of the Office of Weights and Measures, and later with the establishment of the Bureau of Standards, it was not until World War I that the government's stake in standards was really brought home to the nation. In 1917, product diversity was so great that it threatened to hinder the war effort. To deal with the problem, the government set up a Commercial Economy Board of the Council of National Defense. The board's task was to simplify the use of labor, capital, and equipment for all industries. In 1918, the board was incorporated within the War Industries Board, which eventually supervised the manu-

> **The first American standards organizations generally emerged to deal with specific needs as they arose, and thus took a variety of forms.**

facture of over 30,000 articles of commerce.[8]

Concern about the postwar economy led to continued government interest in standards in the period following the war. The hope that wartime simplification efforts would endure was dashed when manufacturers sought to revive consumer demand by increasing product diversity during the "buyers' strike" of 1919-1920. The government's response to the postwar slump was quite the opposite. Inspired by the report, *Waste in Industry*, written by the American Academy of the Federated American Engineering Societies, the government hoped to revive the economy by increasing economic efficiency through greater standardization.[9]

The driving force behind this crusade was Herbert Hoover, the Secretary of Commerce under President Harding. In contrast to the wartime simplification program that had focused on military products, Hoover's program was directed at the economy as a whole. To carry out the program, he organized agencies within the Department of Commerce to provide standards assistance to businesses at their request.

Balancing interests

As more and more stakeholders became involved in standards, it became necessary to differentiate the responsibilities among them. Of prime importance was the relationship between the public and private sectors. Although the government actively promoted standardization at the turn of the century, it gradually relinquished this responsibility to the private standards development organizations. This division of labor continues to this day.

This American preference for private, pluralist solutions is as old as the Constitution itself. Presaging the loosely organized and fragmented standards system to be found in the US, Publius (a.k.a. James Madison), in the *Federalist Papers* (no. 10) argued that the only way to guard against domination by a majority faction is to promote a large number of diverse competing ones. Writing to Thomas Jefferson, Madison summed up this view:[10]

> *Divide et impera*, the reprobated axiom of tyranny is, under certain qualifications, the only policy by which a republic can be administered on just principles.

The founding fathers were successful in framing the Constitution to have just such an effect. From the outset of the new republic, Americans proved to have a penchant for joining factions and establishing associations.[11] Thus support for voluntary, private sector associations was reinforced by a general suspicion of the state and preferences for market solutions.[12] Although these values were often supported more by rhetoric than practice, they were greatly popularized by the progressive movement, which had its heyday in the late 1880s just at the moment when industrialization was primed to take off. Thus, whereas in many other countries government actively sponsored the growth and development of business, in the US industrial development was managed, directed, and financed primarily by the private sector.[13]

The first American standards organizations were in keeping with this tradition. They generally emerged to deal with specific needs as they arose, and thus took a variety of forms. Often established on an industry-by-industry basis, there was little interaction between them.[7] The first American standards organization was the US Pharmacopoeial Convention, which was set up in 1829 to establish uniform standards for drugs. The American Iron and Steel Institute, established in 1855, was the first trade association to develop standards. The American Society of Civil Engineers formed in 1852, was the first scientific and technical society involved in standards development.

The private sector approach survived the war time simplification effort, and was reconfirmed by Secretary of Commerce Hoover when he undertook the standardization crusade in 1921. Hoover was a staunch believer in the private sector. Accordingly, he set up the Division of Simplified Practice in the Department of Commerce to supply guidance, information, and assistance. But compliance with the program was purely on a voluntary basis.[8]

The depression capped the voluntary approach to standards setting. In 1933, Congress cut the Bureau's standards appropriations and impounded its funds. As a result, the staff of the Simplified Practice Division was cut from 40 to four, and much of its work in the area of standards was transferred to the private sector organization, the American Standards Association.

Notwithstanding the American preference for voluntary standards, there were always a number of tensions in the standards-setting community. Consumers were among the first groups to question the system. In the wake of Hoover's standardization crusade, they began to question whether they had derived any benefits from it. It was clear that standardization had saved industry money, but consumers saw little evidence that these benefits were being passed down to them. They also looked to the Bureau for consumer product infor-

mation, an area that business was loath to have government become involved in.[8]

The business community also began to register complaints about the expansion of the Bureau's role, charging it with meddling in its affairs. Alarmed at the establishment of a trade standardization division at the Bureau, the American Engineering Standards Committee (AESC) formally petitioned the Bureau to withdraw from all commercial standards activities. Members of the Bureau refused to attend private sector meetings in protest.[8]

With the government's retreat from the standards arena together with the proliferation of standards organizations, the need for national coordination of standards activity soon became apparent. Standards organizations were not only competing with one another to write standards, they were also writing conflicting standards, thus defeating the purpose.[14]

The first steps toward coordination took place in 1918, during the war, when five national engineering societies, together with the US Departments of War, Navy, and Commerce, formed the nucleus of an organization that was to become the AESC. In 1927, the representatives of 365 national organizations—technical, industrial, and governmental—were officially accredited by the AESC. The following year, this group was reconstituted to form the American Standards Association (ASA). However, despite the ASA, coordination continued to prove difficult because of competition among standards organizations.[14]

The Second World War placed even greater demands for coordination on the US standards community, again raising the question of the government's role in standards. To meet the needs of war, the government became involved in setting standards for consumer goods. At the behest of the Department of Commerce, a special consultant, Carroll L. Wilson, was asked to report on the standards problem, with particular attention to the role the National Bureau of Standards should play in the postwar period. Wilson concluded that both the government and the private sector standards programs fell short. Acting on Wilson's recommendations, the ASA broadened the scope of its concerns to include consumer goods. The ASA constitution was also revised so that all groups with an interest in a particular standard would have a voice in its development. Moreover, the revised constitution required that three members-at-large be included on the association's board of directors to provide a greater voice for consumer interests.[14]

The broadening of the ASA's mandate had only a marginal effect on its ability to serve as coordinator of all private sector standards activities. In February 1965, Francis L. LaQue, vice president of the International Nickel Co., issued a report on the state of the US standards system, which had been undertaken at the request of Herbert Holloman, Assistant Secretary of Commerce for Science and Technology. According to the report, the principle standardization problem in the US con-

With the government's retreat from the standards arena and the proliferation of standards organizations, the need for national coordination of standards activity soon became apparent.

tinued to be that of achieving legitimacy and coordination. The study noted that only 2,300 of the 13,675 nationally produced and used standards were designated as American standards through the ASA. To overcome this problem, the report called for a national coordinating institution for voluntary standardization with international recognition such as that granted other national standards bodies. To assure such recognition, LaQue proposed that this institution have a federal charter and that its standards be officially designated as US standards.[15]

Hoping to gain such a charter, the ASA adopted a new constitution and bylaws and took on the name of the United States of America Standards Institute. Characterizing itself as a federation of trade and other organizations, it redefined its mission. Acting purely as a coordinating body, the Institute no longer intended to develop standards; rather it would orchestrate their development through the combined technical talent and expertise of its member bodies and certify that these standards development bodies adhered to the consensus process.[14]

The government and other members of the standards community resisted the effort of the ASA to strengthen its role. A national charter was not forthcoming, and the FTC protested the use of the name USASI on the grounds that it suggested that the ASA was an official organization of the federal government. A compromise was reached, and the ASA became the American National Standards Institute (ANSI). Reporting on the state of the US standards process several years later, the Stanford Research Institute (SRI) saw little hope for the future. The situation, according to SRI, was in fact deteriorating.[16]

There is little hope that the situation will improve in the next several years. In fact fragmentation is becoming worse. Up through the mid-1960s, a favorable solution appeared possible under the guise of the

> *Responding to consumer concerns and allegations of antitrust infringements and unfairness, the FTC undertook a major investigation of the US standards system.*

quasi-official American National Standards Institute (ANSI).... Reportedly, however ANSI now has less support and less probability of succeeding as the nominal national voluntary standards coordinating agency than it did a decade ago.

At the same time, other standards organizations are attempting to strengthen their individual positions, portending less opportunity for a coordinated effort. A leadership conflict exists and will probably persist for some time.

Consumers and regulatory standards

The federal government's interest in standards was rekindled in the late 1960s and early 1970s in response to consumer concerns about safety and antitrust matters. Ralph Nader first raised the issue in 1965, when he published *Unsafe at Any Speed*, which severely criticized automobile standards as they had been developed by the Society for Automotive Engineers. Other horror stories about the standards system abounded.[17]

Congress was quick to react. In 1967 it set up a National Commission on Product Safety to analyze the effectiveness of consumer product standards. After reviewing more than 1,000 standards, the Commission concluded that the system was "chronically inadequate both in scope and permissible levels of risk."[17, p. 1372] Moreover, it suggested that the voluntary sector process was unable to produce adequate standards, given the dominant role of industry. This attitude was reflected in much of the health and safety legislation that followed, which often made special provision for standards. It was also the basis on which Senator James Aboaurezk, in March 1975, and again in 1977, introduced the Voluntary Standards and Accreditation Act designed to give the federal government considerable control over the voluntary standards system.

Responding to consumer concerns and allegations of antitrust infringements and unfairness, the Federal Trade Commission also undertook a major investigation of the US standards system. After extensive hearings, at which over 200 people testified, it too concluded that the entire standards process should be regulated. It proposed a rule that would require standards setters to meet a substantive "fairness" criterion.[17]

Another outcome of this period was a major increase in the number of federal agencies issuing standards. From the late 1960s until the early 1970s a rash of environmental, health, and safety legislation was passed, and agencies were created to administer these laws. Included among these, for example, were the Consumer Product Safety Commission, the Environmental Protection Agency, and the Occupational, Safety and Health Administration.

Standards development today

Were Publius to observe the US standards process today, he might well be pleased. American standards organizations continue to operate in a pluralistic framework. Almost half of all standards are set as part of a voluntary consensus process, in which all, or most of the key players—including government agencies—participate.

On the other hand, times have changed. The US is no longer an isolated, homogeneous agricultural society in which the greatest danger is rule by an oppressive majority. Quite the contrary. Among the dangers that the US faces today is a loss of competitiveness, due partially to a failure to lead in the international standards development process. Thus, like many reports on the US standards process, Publius might be alarmed by the lack of leadership and failure to develop a national standards policy. However, leadership would require either that the private sector work cooperatively, or that the federal government assume a greater role. Ironically, neither remedy is likely, precisely because of the intensity of conflict that Publius prescribed.

Within the US standards community, there are approximately 400 organizations involved in standards development. These groups are organized and function independently of one another. There are essentially five different types: trade associations, professional societies, general membership organizations, third-party certifiers, and consortia. All of these organizations are private sector, voluntary organizations; they arrive at decisions through a process of consensus, and all have mechanisms for participation, comment, and appeal.[18]

While functioning independently, many of these standards bodies coordinate their activities through the American National Standards Institute. Having no official charter, ANSI is, in effect, the "self-designated" national coordinating body for US standards development organizations as well as the internationally accepted member body in the International Standards Organization (ISO) and the International Electrotechnical Commission (IEC). Receiving the bulk of its financial support from private sector contributions, ANSI's existence depends on its ability to continually meet the needs of its diverse

memberships, a task that has not always been easy.

The voluntary consensus process requires cooperation and trust to succeed. There is little bureaucratic structure to otherwise hold it together. Unresolved disputes and disagreements not only distract from the main purposes of standards setting; they also undermine the legitimacy of the system, both in the opinion of its members as well as in the eyes of the rest of the world. Such is the case in the US standards world today.

Support in the US for private sector standards development hides some deep-seated divisions within the standards community itself. Although most members firmly believe in the voluntary consensus process, they differ about what "openness" means. The American Society for Testing Materials (ASTM) insists that true consensus requires the participation of *all* interested parties, even if this requires subsidizing some groups. On the other hand, ANSI and others argue that due process requires only that the process be open so all have an opportunity to participate. They contend that willingness to pay is an essential measure of interest in the process.

Members of the standards community also disagree about which organizations produce the "best" standards. For instance, many professional societies claim that their standards are technologically superior, since their members participate not as representatives of any group or interest but as individual engineers. Some industry groups argue the opposite. Standards set by professional societies, they contend, do not reflect market forces, and they are often insensitive to industry competitive issues.

Standards-setting bodies also compete to sell standards, which is another important source of contention. Many of these organizations resemble publishers; they orchestrate standards setting in exchange for the right to sell standards and other value-added, standards-related services. Sales from standards, for example, account for 80 percent of the income of ASTM, 60 percent of that of the National Fire Protection Association, and 28 percent of that of ANSI. Competition and turf battles among these and other standard setting bodies often revolve around these sales. These struggles are likely to become even more intense and convoluted in the future with the growth of a world market for standards and the emergence of new global competitors. This economic competition is compounded by personality conflicts in the standards setting community, some dating back a number of years. There is little trust or respect among the leadership. People characterize one another in acrimonious terms.

The interests of some standards-setting organizations are also beginning to diverge from those of manufacturers. In a highly competitive global economy, for example, it is important for manufacturers to have their standards adopted on an international basis. They may even want to "give" their standards away in an effort to develop new markets. However, such a policy is not in the interest of those standards-setting

Among the dangers that the US faces today is a loss of competitiveness, due partially to a failure to lead in the international standards development process.

organizations, whose livelihoods generally depend on standards sales. In addition, manufacturers may want to speed up standards development and implementation, but standards-setting organizations often hesitate to put their standards electronically on line due to copyright concerns.

Conflicts in the standards community weaken the US position internationally. Aware of these disputes in their most minute detail, European standards makers use them to their advantage. Even so, Europeans would prefer that the US presented a united front to the rest of the world. "The US," they say, "is a major economic power, and it must play its role in international standard setting accordingly." Europeans emphasize how difficult it is to negotiate with one body speaking authoritatively for the US, "when you are unclear about its actual power, and who it really represents." They complain that one moment they are told ANSI speaks for all the US; but the next, ASTM is knocking at their doors.[18,p.13]

Internecine warfare in the standards community also raises questions about the ability of the voluntary standards organizations to carry out the public trust delegated to them. In a recent display of these problems, ANSI charged before the Office of Management and Budget (OMB) that certain parties in the Department of Commerce are undermining ANSI's authority through their actions. However, three other major US standards-setting organizations quickly took exception to this charge, claiming that they fully support the Department of Commerce's actions.[18]

Paralleling the lack of unity in the private sector standards community is a lack of coordination and policymaking at the federal level. While this is not a new problem, its consequences will be more serious in the future. As the US expands its role in a global economy, new trade-offs among standards goals must be negotiated. Free-trade objectives are already coming into conflict with environmental and safety goals. Under such circumstances, coordination and conflict resolution among federal agencies are essential. Moreover, with the growing importance of standards, rapid technological advance, and the shift to a global economy, the federal

Any politically viable solution will entail compromises and a fresh perspective.

government needs some ongoing organizational capability to identify problems, set goals, and evaluate system performance.

A 1977 Department of Commerce report on the US standards-setting process, as well as the 1965 LeQue report, called for a unified, national standards policy. They proposed the establishment of some form of government body, where policies should be considered. However, this type of solution was unpopular—especially in the business community—and nothing came of it.

The problem of coordination was eventually addressed on a limited scale with the establishment of an interagency committee. In accordance with OMB Circular A-119, the Department of Commerce (DOC) was directed to set up an interagency consultative mechanism to advise the Secretary and agency heads in implementing federal standards policy (as defined in the Circular); to coordinate agency views; and to develop, where possible a single, unified position. DOC assigned this task to the Interagency Committee on Standards Policy, which operates under the direction of NIST. Overall oversight rests with OMB, and the committee is required to report back to it every three years.

While active during its first year, as of 1992 this interagency committee had reportedly not met for a year and a half. Meetings focused on implementing the federal policy to encourage agency use of voluntary standards, as directed in its mandate. The committee also set standards for agency participation in voluntary standards bodies and laid out guidelines for public sector use of private certification bodies. Participants claim, however, that scant attention was devoted to evaluating existing policy or finding ways to improve it. Nor was there much effort to identify future standards issues or to view them strategically as part of the industrial infrastructure.

The Office of Management and Budget reviews the work of the Interagency Committee every three years. Although OMB is the ultimate coordinating mechanism in the federal government, it can do little more than establish a policy directive. There is little staff support in the area of standards. The deputy director of the Office of Federal Procurement Policy is in charge of overseeing Circular A-119. However, there is no one person at OMB who focuses explicitly on standards.

Having no comprehensive national standards policy of its own, the US has tended to disregard or underestimate other governments' efforts to build standards into their industrial policies. Most other countries not only view standards as a strategic marketing device to help develop markets abroad but also as part of their industrial infrastructure, to enhance economic productivity, reduce costs, and provide for greater quality control. Thus, for example, the European Community as well as Japan have programs to educate and train domestic companies in the use of standards, to subsidize national participation in international standards organizations, and to subsidize and provide technical training for standardization efforts in developing countries where there is considerable market potential.

The US has no equivalent policies or programs. Failure to appreciate the implications of standards policies and the growing importance of standardization could have serious consequences for US industry. As the US adjusts to a changing global economy, more and more industries are not only dependent on trade, they are also increasingly affected by standards. It is estimated, for example, that of $83 billion in exports of manufactured goods, some $40 billion is, or will be, subject to European Community product safety standards alone.

Standards setting is likely to be even more important to the nation in the future due to the economy's growing reliance on technology. Just as specialization and assembly line production provided an impetus for standardization during the industrial era, so too networked production and computer-assisted work are increasing the demand for standards today. Machines require more precision than humans, because they are less flexible in adjusting to errors and omissions. Moreover, in a global information-based economy, networking technologies provide a basis for productivity and economic growth. These technologies will become the basis of an infrastructure for all economic activity. If networks fail to interconnect for lack of standards, the nation could suffer considerable economic loss.

The growing pace of technological change will also drive the need for standards development. The faster the advance of technology, the greater the risk in R&D and product development. Standards-setting processes can help to reduce uncertainty in a rapidly changing technological environment. Participants in the process learn first hand about new technologies. Moreover, by developing reference models in the anticipation of actual standards, manufacturers have a general target toward which they can direct technology development. Standards setting, therefore, will increasingly be an important aspect of any national economic policy aimed at encouraging innovation and economic growth.

Many of the standards-setting problems just identified are persistent problems, which have been cited before. The inability of the US to deal with these problems reflects the high stakes and significant ideological differences involved. There

are no perfect solutions. Stakeholders strongly disagree about what constitutes a perfect state of affairs. Thus any politically viable solution will entail compromises. Above all, it will require a fresh perspective that objectively considers both the problems of the system and the ways in which all participants—public as well as private—can join to resolve them. ▣

Acknowledgments

References

1. *The Growth of the American Economy*, H. Williamson, ed., Prentice-Hall, New York, 1951, p. 721.
2. E. Kirkland, *Industry Comes of Age: Business, Labor, and Public Policy*, Holt, Rinehart, and Winston, New York, 1961, pp. 1329-1483.
3. J. Beniger, *The Control Revolution: Technology and the Economic Origins of the Information Society*, Harvard Univ. Press, Cambridge, Mass., 1986, pp. 15-16.
4. G.S. Radford, *The Control of Quality in Manufacturing*, The Ronald Press, New York, 1922, p. 270.
5. A. Edwards, "Standardization in the Household," *Notes from the Annals: Standards in Industry*, The American Academy of Political and Social Science, 1928, p. 25.
6. A. Nesmith, "A Long, Arduous March Towards Standardization," *Smithsonian Magazine*, Feb. 1985, p. 185.
7. C. Cargill, *Information Technology Standardization: Theory, Process and Organization*, Digital Press, Bedford, Mass., 1989, p. 21.
8. R.C. Cochrane, *Measures for Progress: A History of the National Bureau of Standards*, National Bureau of Standards, Washington, D.C., 1966, p. 167.
9. R.M. Hudson, "Organized Effort in Simplification," *Notes from the Annals: Standards in Industry*, The American Academy of Political and Social Science, 1928, p. 25.
10. M.F. Plattner, "American Democracy and the Acquisitive Spirit," *How Capitalist is the Constitution?*," R.A. Goldwin and W. Sebambu, eds., The American Enterprise Institute, Washington, D.C., 1982, p. 11.
11. *The Voluntary Sector in Comparative Perspective*, R. Wuthnow, ed., Princeton Univ. Press, Princeton, N.J., 1991.
12. W.P. Adams, "Republicanism" in Encyclopedia of Political History, J.P. Green, ed., Scribner's, New York, 1984.
13. D. Vogel, "Government-Industry Relations in the United States: An Overview," *Comparative Government-Industry Relations*, S. Wilks and M. Wright, eds., Clarendon Press, Oxford, UK, 1987.
14. Congressional Research Service (CRS), Science Policy Division, *Voluntary Industry Standards in the United States: An Overview of Their Evolution and Significance for Congress* (Report to the Subcommittee on Science, Research and Development of the Committee on Science and Astronautics), US House of Representatives, 93rd Cong., 2nd Sess., July, 1974.
15. F. LaQue, *Report of the Panel on Engineering and Commodity Standards of the Commerce Technical Advisory Boards, Parts A and B*, Washington, D.C., 1965.
16. *Industrial Standards*, Stanford Research Institute, Menlo Park, Calif., 1971, p. 3.
17. R. Hamilton, "The Role of Nongovernment Standards in the Development of Mandatory Standards Affecting Safety and Health," *Texas Law Review*, 56, 1978.
18. Office of Technology Assessment, US Congress, *Global Standards: Building Blocks for the Future*, Report TCT-512, US Government Printing Office, Washington, D.C., Mar. 1992.

Linda Garcia is a project director and senior associate at the Office of Technology Assessment, a research arm of the US Congress. As a member of the Telecommunications and Computer Technologies Program, she has worked in the areas of information technology and education, human resources for information technology R&D, and information policy. She served as project director for the OTA assessments *Intellectual Property Rights in an Age of Electronics and Information, Critical Connections: Communication for the Future, Rural America at the Crossroads: Networking for the Future,* and *Global Standards: Building Blocks for the Future.* She presently heads up a study, *The Electronic Enterprise,* which examines how telecommunications and computer technologies will likely affect American business.

Garcia holds masters degrees from the Department of Political Science and the School of International Affairs at New York's Columbia University.

Direct questions to the author at Congress of the United States, Office of Technology Assessment, Washington, DC 20510-8025; or lgarcia@ota.gov.

Reader Interest Survey

Indicate your interest in this article by circling the appropriate number on the Reader Service Card.

Low 159   Medium 160   High 161

Markets, Standards, and the Information Infrastructure

Standards development plays a powerful role in shaping important contemporary events and in charting the course for tomorrow's information infrastructure. Recognizing the dual nature of standards—as both coordinator and constraint—clarifies this process. While the constraints they impose may be costly in the short run, standards ultimately coordinate technical development on a broad scale. Progressive decentralization in their development has unleashed massive entrepreneurial activity, fostering dynamic economic growth around the world.

Shane Greenstein

University of Illinois,
Urbana-Champaign

Today's worldwide information infrastructure encompasses a broad spectrum of activities. Telephones, local area networks, wide area networks, and supermarket scanners all play a part. That infrastructure involves both simple and sophisticated equipment—everything from wireless communication devices, to microprocessors, to thousands of miles of copper cables.

This information infrastructure did not arise overnight, nor did any single policy vision guide it. A host of legal, economic, and historical factors shaped its development over most of this century. The processes were decentralized, usually market oriented, and seemingly too chaotic for any organization to control. Of course, some have tried. AT&T, IBM, the Federal Communications Commission, and the US Department of Defense, among others, have briefly, and sometimes successfully, coordinated the development of one component of the whole.

No single policy vision coordinates infrastructure development today. Nor could any centralized decision process possibly guide such a complex engineering network. By default, decentralized market mechanisms, private firms, and standards development organizations take responsibility for many technical standards within the information infrastructure.

Approaching the standardization process. Standardization activity plays a dual role when decentralized, market-oriented decision-makers make technical decisions. These dual roles, as coordinator and constraint, show up in both short- and long-run analysis. Failure to account for both roles has frequently obscured our understanding of the standardization process.

In the short run, when market structure is relatively stable, standards coordinate contemporary and anticipated market behavior. The costs of using a network of components usually decline because standardization reduces the interconnection expense. Standards let component designers anticipate interconnection requirements so they can improve their part of the system. With standards, system users can invest in assets and rest assured that loss of connectivity will not depreciate the asset's value.

Yet this coordination benefit is not free. Standards limit the choices of users and vendors alike. Both become locked in to a set of technical constraints that they may change only at a high cost. Moreover, vendors recognize the strategic importance of locking users to a standard and spend, or waste, considerable resources on manipulating its development.

In the long run, the ultimate importance of standardization arises from its impact on technical change. Because many parts of the information infrastructure have not reached the stasis associated with mature product markets, standards shape technology-based decisions. Stan-

dardization issues lie at the core of developments in digital cellular telephones, high-definition television, and large LAN communication protocols.

In this setting standardization also plays a dual role as constraint and as coordinator. Stable, functional, predictable standards coordinate technical development. Yet standards also lock in users and suppliers over the long haul. Lock-in is especially costly when technical possibilities change rapidly, removing previous technical constraints and imposing costly new ones. Though standards will constrain technical improvement, that development will occur sooner and will involve more development among more components.

Definitions and distinctions. The difficulty with any analysis is that today's information infrastructure is a "network of networks." It consists of a hodgepodge of public and private telephone networks, private local and wide area networks, mainframe- and mini-computing centers, and numerous communication bridges between various subnetworks. Telephone companies, computer hardware and software companies, satellite operators, governments, and virtually every user play some role in this network.

Analysis breaks through this hodgepodge by focusing on one "economic network" at a time. All buyers and suppliers who have economic incentive to care about a system's technical features comprise such an economic network. Either all users desire to communicate with one another, as in a traditional telephone network, or all users need electronic components to work with one another, as when an industry wide network of buyers uses the same standard bundle, or the minimal set of components necessary to ensure system performance.

Notice that the use of "network" here is not conventional. Economists view telecommunications networks as more than just their physical linkages and electronic signals, more than just the physical equipment extant at any given time. Economic relationships extend beyond physical boundaries of equipment. Though many buyers and sellers of the same information technology may not buy equipment or services from the exact same supplier, they still may be a subset of the same economic network if they use compatible equipment.

All activity in an economic network centers around interoperability—whether a component may serve as a subsystem within a larger arrangement of components. In the simplest case, compatibility standards define the physical fit of two components. Familiar examples are modular phone jacks on telephone cords and handsets, and compatible telephone switches. More complex are the standards that determine electronic communication channels. The need for these standards is obvious, since successfully filtering, transmitting, and translating signals across telecommunication networks requires precise engineering. Similar needs arise in the design of circuitry linking computers, their operating sys-

Standardization issues lie at the core of developments in digital cellular telephones, high-definition television, and large LAN communication protocols.

tems, and application software programs.

More generally, compatibility solves but one issue in a wider array of coordination problems. Most on-line commercial networks—Prodigy, Compuserve, America Online, or the private networks of thousands of commercial organizations and private firms—are sophisticated electronic networks. These often involve on-line transaction processing, employ a mix of sophisticated telecommunications and computing equipment, and must operate reliably on a daily basis.

Accomplishing these various functions involves all the coordination activities associated with the successful management of a business enterprise. Products and services must be defined and tied to billing. Output must be controlled and its quality assured. Electronic signals must be routed without hesitation. An organization must also develop capital capacity and plan the requisite staffing to meet long-run service needs. Sometimes these decisions involve coordinating actions within a single organization. Often they involve coordinating decisions across divisions within the same company, or among upstream and downstream vendors, or between a vendor and a governmental regulator.

Economic research to date focuses primarily on the factors influencing the development of compatibility standards. This focus on the nexus of economics and technology is a bit narrow, since it virtually ignores the important organizational costs just mentioned. Nonetheless, since interoperability is necessary for coordination on any level, this restricted view does not invalidate the merits of the analysis of compatibility. It simply means that typical analysis ignores lots of the messy details of coordinating organizations in practice. As I will point out, sometimes this hole matters and sometimes it does not.

Another key is the economist's taxonomy of processes that develop standards. Unfettered market processes may develop standards as a de facto result of either a sponsored or an unsponsored market process. In a sponsored process, one or more entities, suppliers, or cooperative ventures create inducements for other economic decision-makers to adopt a particular set of technical specifications and become part

Table 1. Short run analysis: trade-offs between different market structures.			
Criteria	Unsponsored standard	Dueling sponsors	Single-sponsored standard
Decision-making	Diffused to many firms	Concentrated in a few firms	Concentrated in a single firm
Severity of coordination problem	Difficult to reach agreement between all interested vendors and users	Depends on willingness of vendors to design components that mix and match	All decisions internalized by single firm—depends on management of firm
Pricing	Typically very competitive—pricing close to cost	Oligopolistic pricing—typically some markup over cost	Monopolistic pricing—high markup over cost
Primary distortion	Decisions subject to bandwagons—society will likely not get optimal technology	Vendors' strategy determine networks—vendors will lock-in users and lock out rivals	Monopolist will manipulate technology to own advantage—blockade as much entry as possible

of an economic network (such as predivestiture AT&T-sponsored telecommunication standards). An unsponsored process has no identified originator with a proprietary interest, yet follows well-documented specifications (the QWERTY keyboard). Voluntary industry self-regulation may also play a role when economic networks arise out of the deliberations of standards development organizations. Of course, government bodies may also shape the development of economic networks (such as the FCC).

Government organizations have no compelling reason to involve themselves in the development of every network. They often do so because important public policy issues are at stake, as when domestic and foreign firms use standardization as a competitive weapon. They often do not do so because external forces, such as dramatic technical change, outstrip the ability of any administrative process to guide events, making it easier to leave decisions to market participants. When to rely on a market process instead of on government decision-making is an open and active topic of debate, one that usually hinges on trade-offs between imperfect market processes and imperfect government intervention.

Short-run analysis

Short- and long-run analyses require different approaches. Short-run analysis presumes that the number of key decision-makers—such as firms or potential users—is virtually fixed. This is not bad if many rigidities—a firm's technical expertise, economies of scale, and various other competitive advantages associated with incumbency—limit how many firms can feasibly produce for a market in the short run. By implication, short-run analysis is not appropriate for investigating how technical innovation influences the adoption of standards and the number of suppliers, and vice versa. Also, since rigidities differ in importance in different markets, the appropriateness of this type of analysis will also differ by industry.

For short-run analysis, we conveniently distinguish between networks in which many suppliers provide related services, a few do, or only one dominates. Table 1 summarizes some of the differences that arise between such systems. These distinctions help organize insights about patterns of outcomes and the factors that produce them. Of course, determining which markets belong in which categories is not always obvious in practice. Indeed, much controversy is essentially argument over which type of analysis applies to which specific market.

Many decision-makers and too many cooks. Standardization may not easily arise when decision-making in a market is diffuse—when a market has many buyers and many sellers, none of whom is responsible for a large percentage of economic activity. This trend is disturbing since diffuse market structures are typically very competitive and tend to allocate scarce resources efficiently through price mechanisms. Many policy issues would be simplified if diffuse market structures gave rise to desirable standards.

Coordination problems and the lack of sponsorship. When decision-making is diffuse, coordination problems sometimes arise. Such terminology is not a statement about whether an economic enterprise coordinates its own employees around a single objective. Rather, it means that all potential users

Table 2. Standardization and technical change: trade-offs between different market structures.		
Criteria	Single-sponsored standard	Unsponsored standard
Systematic innovation	All decisions internalized within single firm—likely to be accomplished as fast as technically feasible	Firms must coordinate changes within SDOs—likely to be administratively difficult and slow
Component innovation	Sponsor tends to resist cannibalizing rents on existing products—component innovation is slow	Component vendors must frequently innovate to stay ahead of the competition—component innovation is fast
Coordination of technical change	Firm's administrative process coordinates changes in the design of own products	No one is responsible for technical change—is uncoordinated and uncertain
Degree of lock-in in the long run	Likely to be high because users have no alternative	Lock-in is as low as possible because competing vendors will try to keep lock-in low

and suppliers could benefit from as much technical interoperability as possible, but instead go off on their own. The sheer number of decision-makers hinders adequate communication that would solve the coordination problems. Even if all firms could communicate, differences of opinion make consensus unlikely.

Moreover, standards that serve as focal points are unlikely to arise very easily, because every potential supplier and user of a standard is a small part of the whole. Each decision-maker has too little incentive to make the investments that will coordinate the design decisions of other users and lead to general interoperability. Market structure alone may hinder network growth because standardization does not arise, or it arises too late. The slightly different Unix systems proliferating in the 1970s and 80s serves as a good example.[1]

This observation immediately leads to one disturbing prediction for the growth of private telecommunications networks: if standards are unsponsored, different firms' networks will likely not work with one another without considerable adjustment. Private networks often develop according to internal imperatives. When these networks grow larger and brush up against one another, they may be unable to work together simply because no sponsor ensured that they were initially developed in a technically compatible manner. For example, after the introduction of supermarket scanners, suppliers took years to coordinate their deliveries with the inventory management of grocery stores, if they coordinated them at all. Similar factors have slowed the introduction of scanners into the retail clothing sector. Table 2 summarizes some of the differences between sponsored and unsponsored standards structures.

When unsponsored economic networks develop and build capacity, they often swell and shrink for many reasons that may have only a tenuous connection to the long-term economic welfare of market participants. In a characteristic bandwagoning effect, networks may be slow to start when they are small. Many potential adopters will sit on the fence, waiting to make expensive and unrecoverable investments until a large fraction of other users choose a clear technical standard. Networks may not develop at all if most participants are lukewarm about a new standard due to technical uncertainty, for example, even though all would collectively benefit from it. Alternatively, bandwagons may also gather speed (remarkably!) quickly once a network becomes large enough to justify investments by potential adopters who, in the early phase of development, had delayed making commitments. The lack of communication between all the potentially affected decision-makers exacerbates such bandwagons, though professional organizations can often provide communication channels to bridge some of the troubles.

Lock-in. A costly problem arises if most vendor and user capacity for a network becomes locked-in to a technical alternative, making it costly for users and suppliers to change fundamental technical specifications. Either hardware or software embodies difficult-to-change technical features, or humans cannot be retrained easily to work with a different technology. These costs are especially high when a network must change—be upgraded, expanded, or replaced—and the network serves as an essential part of an organization's day-to-day operations. Change risks significant downtime that arises from the costs of fixing the almost inevitable mistakes any change produces. In the 1980s, the Federal Aviation Administration updated its air traffic control systems

QWERTY keyboard

The best known example of past decisions constraining future choices is the QWERTY keyboard, which was explicitly designed to slow down the typists of the 19th century. David[1] argues that the interaction of uncoordinated decisions by typing schools, typewriter manufacturers, and early typists resulted in the adoption of the QWERTY keyboard and its persistence past it useful life. A superior alternative exists, yet market participants have never coordinated a switch.

A more contemporary example of the lock-in due to intertemporal links is the MS-DOS operating system for PCs. Because it was designed for 8-bit microprocessors, it uses the available RAM on today's 32-bit microprocessors poorly. Similar examples from information technology markets, if perhaps less dramatic than QWERTY, include the development of AM stereo, FM stereo, microprocessor designs, and Unix operating systems.

1. P.A. David, "Clio and the Economics of QWERTY," *Am. Economic Rev.*, Vol. 75, May 1985, pp. 332-336.

across the country; the small margin for error made the upgrade especially difficult.

Lock-in produces two related problems:

- A network may not become as large or as valuable as possible because users lock in to a disparate variety of formats and each finds it costly to change later.
- If many potential adopters wait for a "shake-out," early adopters may make crucial choices between technologies, thus bearing a disproportionate influence over standards. Technical designs may not permit easy alteration to accommodate the different needs of the later decision-makers.

Perhaps the disproportionate influence of early users is justified because these same users bear a high risk for being intrepid—their investments in a network can become obsolete or "orphaned." However, this argument sidesteps the question of whether society gets an optimal technology or not, which is the central policy concern.

The issue is not solely that past investment influences future technical choices, which happens quite often and complicates choices, but is a sober fact of life. For example, the installed base of color television sets in the US today use a set of standards that is incompatible with many of the new HDTV standards. Many observers think that abandoning this installed base is too costly and, thus, recommend using a

high-definition standard that is backward compatible with the installed base, even if doing so sacrifices some of the pictorial quality possible with HDTV technologies or raises costs.

More importantly, society can be locked-in to the wrong technology after the fact. When viewed with hindsight, "society" could regret previous decisions. Even though past choices constrain future choices, future decision-makers never have an opportunity to persuade previous decision-makers about that choice. Hence, past choices will likely be short-sighted. The QWERTY box elaborates on several ways past decisions have constrained future choices to society's detriment.

Can a small number of cooks do any better? Diffuse decision-making leads to situations where 1) communication and sponsorship are unlikely, and 2) coordination problems are likely. Therefore, market structures with few vendors then should not suffer as much from coordination problems. However, such a conclusion is hasty if not qualified properly. In markets with few vendors, the proprietary interests of the vendors lead them to take strategic actions designed to produce outcomes they favor. While this reduces the severity of some types of coordination problems, it also induces other types of distortions.

The "dueling sponsors" arrangement best illustrates these concerns. Here each sponsor has proprietary interests in an array of components that perform similar functions, but competitors employ different technical standards. The VHS/Betamax duel in the VCR markets is a well-known case. Such battles are common today in high-tech industries (IBM vs. DEC in minicomputers, MS Word vs. Word Perfect in word processing, FDDI vs. ATM in network communications). These duels may start as multifirm contests but quickly reduce to a handful of dominant participants. Sometimes a fringe of niche market suppliers follows the leaders, leaving two or three technical standards to dominate all choices.

Network duels also commonly arise as subplots to related larger product market duels. Various banks may belong to incompatible ATM networks, and United Airlines and American Airlines sponsor competing airline reservation systems. If recent experience is any guide, this type of market structure will likely characterize many, if not most, private economic networks in the future information infrastructure.

Dueling locks-in users. Economists are of two minds about dueling. On the one hand, dueling may prevent the economic network from becoming as large as it possibly could, even if all users would benefit from a larger network. Unlike an unsponsored network, dueling encourages a vendor to lock in buyers. Dueling sponsors have incentives to design incompatible systems if incompatibility raises the costs to users of switching to a rival sponsor's system. Similarly, the sponsor of a system would like nothing better than to raise the costs to the experienced user from switching vendors, since it makes a user reluctant to change networks.

Vendors like to be the exclusive provider of a technology to a locked-in buyer for several reasons:

- It provides the sponsor with market power during any repeat system purchase.
- It guarantees a stream of related business. In computing networks, for example, locked-in buyers will purchase CPU upgrades from their system sponsors, and often a majority of their peripherals and software.
- Locked-in users can be manipulated for competitive advantage. In the case of computer reservation systems, the sponsoring airlines were accused of locking in travel agents and then manipulating the screen to favor the flights of the system sponsor.

Similar factors, as well as several pricing issues, prevented ATM networks from working together as one large network for many years.

Notice that a vendor may desire to lock in its buyers, but a vendor's competitors will desire the opposite. While a vendor may try to raise the cost of switching, a rival may be working equally hard to lower those costs. From society's standpoint much of this activity is wasteful. Wouldn't society be better off if all competitors ignored lock-in and directed all their energy at making better products? Yes, but this will rarely occur because of the strategic importance of standards in a competitive duel.

As with unsponsored economic networks, the market's choice between dueling systems retains the sensitivity to small events, which is some cause for concern. The AC/DC box illuminates this concern with a much researched example from the early history of electrical power supply.

Dueling has its good points. Despite the foregoing, economists are not uniformly pessimistic about dueling, which is where some confusion arises. Sometimes dueling sponsors will not design incompatible systems. When rival sponsors provide components that perform different or complementary functions, compatibility permits many mix-and-match possibilities among the components of rival systems. In turn, the profitability of producing compatible components (despite increases in competition) rises. The market for stereo equipment is a familiar example, as is the market for personal computer hardware clones and software applications under the DOS standard.

Thus, dueling sponsors are likely to find it worthwhile to make investments to reduce interoperability costs when they do not produce every type of component. They might as well if each has comparative advantage in the design and production of some but not all components, a common occurrence when market participants have different technical capabilities. This is probably a good explanation for the willingness of many firms, AT&T and IBM increasingly so, to participate in markets with nonproprietary standards.

AC/DC wars

Though engineering evidence seems to suggest that alternating current is probably superior to direct current for widespread use, many other factors, including beauty contests with pretty girls and the decisions of crucial industry participants, such as Edison and Westinghouse, and the character of the gateways between the two technologies, played a crucial role in the success of AC over DC.[1] To appreciate the lengths to which Edison and Westinghouse went, consider that Edison attempted to persuade New York state to use AC in its electrocutions to foster the perception that DC, which Edison sponsored, was a safer form of electricity. (Incidentally, DC made for more efficient electrocutions, so Edison lost this particular skirmish.)

More recent studies show that the development of the VCR standard was sensitive to the relationship of Sony and Hitatchi corporations, the seemingly minor (and temporary) ability of VHS to record longer, and, most crucially, the timing of the introduction of video cassettes, which occurred unexpectedly and rather randomly from the viewpoint of the major VCR manufacturers.[2]

1. P.A. David and J.A. Bunn, "The Economics of Gateway Technologies and Network Evolution: Lessons from Electrical Supply History," *Information Economics and Policy,* No. 3, 1988, pp. 165-202.
2. M.A. Cusumano, Y. Mylondadis, and R. Rosenbloom, "Strategic Maneuvering and Mass Market Dynamics: The Triumph of VHS over BETA," Working Paper 91-048, Harvard Business School, Cambridge, Mass., 1991.

Dueling standards may also be economically efficient if a variety of standards is appropriate for a variety of potential problems. The crucial question is whether the market will permit entry of a new standard suited to a minority of users. This may depend on the strength of lock-in effects or the success of actions of system sponsors to foreclose or induce entry of complementary products, such as software. Specific conclusions depend on careful analysis of particular industries.

Competition and innovation also counterbalance some of the distortions from lock-in, giving rise to another cause for optimism. Monopoly profits may be dissipated through competitive bidding between the rival system sponsors. Since many buyers anticipate that their vendors will later gain monopoly benefits from exclusive sales of complementary products, they will demand compensation before commit-

> *Intersystem competition was a primary driver of computer system innovation in the 1960s and 1970s. But this benefit sometimes accrues only to new users, not to users with an installed base of equipment, who are already locked in.*

ting to investment in network capacity with proprietary features. Such demands can elicit "promotional pricing" from sponsors.

The good news is that the networks with long-run economic advantages are likely to provide bigger price discounts. Also, competitive bidding for new customers may spur incumbent system vendors to innovate. Some observers argue that intersystem competition was a primary driver of computer system innovation in the 1960s and 1970s.[2] The bad news is that this benefit sometimes accrues only to new users and not necessarily to users with an installed base of equipment, who are already locked-in.

Dueling may also induce actions that ultimately lead to the success of one economic network but also the loss of the sponsor's control over it. A firm may broadly license a technology to establish it as a standard, but in so doing, sacrifice its control over the standard and much of the monopoly profits associated with that control. Sun Microsystems' liberal licensing strategy with the Sparc architecture exhibits some of these features.

Another variant of this phenomenon arises when a firm designs a product that does not contain proprietary technology. A nonproprietary system induces entry of more peripheral and software suppliers and hardware clones. This makes the hardware conforming to the standard more valuable to users, while the entry of more clones reduces the price. The development of software and peripherals for the IBM-compatible PC followed this pattern. Once the standard was widely accepted (partly as a result of all this entry), IBM no longer garnered much of the rents from being the original sponsor of the standard. Indeed, today IBM and a consortium of private firms are battling to determine the direction of the next generations of "IBM-compatible" machines.

Perhaps the greatest weakness of the economic analysis of dueling systems is also its strength: the long list of possible outcomes. Prediction is quite difficult, particularly in view of the multiplicity of pricing and promotional strategies typically available to firms in information technology markets. Translating economic analysis into useful managerial advice pertinent to a specific market can be difficult.[3]

A single chef makes a menu of favorite recipes. Placing a single sponsor in charge of a standard is a natural solution to coordination problems. The structure of a single firm internalizes all design decisions and upgrading and maintenance problems. Unifying control within a single firm generally eliminates competing designers, providing users with certainty about who controls the evolution of standards and their ultimate compatibility. We cannot overemphasize this potential benefit from single-firm sponsorship, especially in markets subject to uncertain and rapid changes in technology. Many readers will recognize this as the traditional model of telephone networks under AT&T's predivestiture leadership and as IBM's vision for integrating computers and telecommunications under the System Network Architecture model. Many other firms have also tried to adopt this model, though competition often forces them into duels.

The problems with dominance. Unfortunately, single-firm sponsorship by a supplier also brings much baggage with it. Generally, large firms have disproportionate influence upon market processes, which they manipulate to their advantage, to the detriment of society's long-term interests. Most of these concerns fall under the realm of anti-trust economics or traditional regulatory economics.

Anti-trust and regulatory issues arise whenever a dominant sponsor competes with small plug-compatible component suppliers in some or all component markets. IBM battled plug-compatible component suppliers from the late 1960s onward. Similarly, from the mid-1950s on (and growing thereafter), AT&T faced competition in customer-premises equipment markets and long-distance. Today the Regional Bell Operating Companies are beginning to face competitive bypass to their services from nonregulated suppliers of fiber-optics. Anti-trust concerns arise because the dominant firm always wishes to prevent the component firms from gaining market share (and may even want to drive them out of business), while society may benefit from the added competition. Controlling and manipulating technical features of a product, or effectively raising the costs of interconnection, may enhance a dominant firm's strategies aimed at gaining competitive advantage.[4]

Essentially, a large system sponsor and small component supplier do not possess the same incentives to be interoperable: a small firm usually wants interoperability and a large firm does not. The benefits to vendors from accessing a rival network's users is counterbalanced by the loss of market power from facing competition from a rival vendor. Vendors with larger markets are less likely to desire compatibility with

smaller rivals (than the smaller rival does with them) because larger firms gain less from selling to a few more customers and could lose more from facing more competition. IBM's role in blocking the development of ASCII standards for mainframe computers[5] and allegedly in plug-compatible equipment markets as well[6] exemplifies this behavior.

Dominance and policy issues. Competitive behavior presents two difficult issues:

- Under what conditions will a dominant firm manipulate a technology to its advantage and to the detriment of potential entrants and consumers?
- Can and should such behavior be regulated? That is, do the benefits from preventing inappropriate market conduct outweigh the side-effects from imposing an imperfect legal or regulatory rule?

Most observers stumble on the first question, and even if observers clearly describe (in nonpolemic tones) a sponsor's strategies that are inappropriate for society, they may fail on the second set of issues. Policy rules that prevent inappropriate behavior will almost always deter perfectly acceptable behavior as well.

As a result, many relevant debates remain unresolved. Such debate, for example, surrounds any analysis of leveraging—using monopoly power in one component market to gain competitive advantage in another. Most economists agree that courts have carelessly applied this concept, though few agree on an appropriate definition. Definitions aside, a network sponsor surely can delay entry of complementary component suppliers, and possibly foreclose entry altogether. AT&T's resistance to designing modular telecommunication connections delayed entry of competition for customer-premises equipment.

However, the unresolved policy question is whether such behavior should or can be regulated to any good end. One big problem, though not the only one, comes when courts get in the business of second-guessing every innovation, especially those with exclusionary features. Such unwarranted interference may have a chilling effect on many firm's willingness to introduce any innovation, which normally is not in society's long-term interest.

The legacy of the IBM antitrust victories has left firms considerable latitude in the use of standardization for strategic purposes. However, future cases will probably further test key legal rulings. The recent Federal Trade Commission investigation of Microsoft and the recent anti-trust suits against Nintendo foreshadow such a trend. Also, important legal standards are likely to come from several ongoing trials that raise issues in intellectual property rights in computer software standards, and also in trials that attempt to modify Judge Green's restrictions on the Regional Bell Operating Companies.

Vendors with larger markets are less likely to desire compatibility because they gain less from selling to a few more customers and could lose more from facing more competition.

In sum, there are biases inherent in having a dominant firm. There are also gains from coordinating product characteristics and standards. No consensus on these issues is likely to emerge soon in telecommunications or any other network industry. Issues regarding sponsorship are likely to remain controversial as long as there is no consensus regarding the proper role for monopolies in nascent industries.

Long-run analysis: Changing the basic recipe

The discussion until now has treated the growth of standards as the byproduct of initial market conditions. Such an approach is obviously incomplete for long-run analysis: as network industries mature, standardization alters a market's structure. While this feedback is easy to recognize, it is not well understood. Usually, several factors may be at work at once in the long run.

Converters: To bridge or not to bridge. Perhaps the most unsatisfying feature of the short-run analysis of economic networks is its use of a strict concept of lock-in. Are some features of a technology more immutable than others? Are there degrees of lock-in? Economic analysis has yet to fully explain situations where interoperability is in constant flux, where the "standard bundle" changes frequently as suppliers update and revise products. The analysis of converters partially addresses this issue.

Converters (or translators or emulators) bridge the gap between otherwise incompatible networks. These products, whether supplied by a system sponsor or third parties, reduce the costs of interoperability. A number of third-party vendors today supply programs that enable Apple Macintosh computers to use IBM software. Also, many software programs now come with simple utilities for translating text or databases from the format of one software package to another. These bridges clearly have value to buyers, so they arise in virtually every economic network.

The interesting feature of converters is that vendors

One can make a case for limited government intervention aimed at guaranteeing a minimal amount of interoperability, at least to induce technical change and capacity investment.

unequally share the costs and benefits from their introduction and refinement. At the very least, the incentives to introduce a converter then will probably not match society's. Most of the benefits of the IBM/Macintosh converter accrue to users of a Macintosh system, so IBM has little incentive to help in its development. Indeed, a public vendor sometimes may actively seek to prevent the entry of gateways and sometimes not, depending on the costs and benefits. Observers often accuse IBM of discouraging global compatibility between all computer languages.[5]

Because converters may lead to large, discreet changes in the boundaries of competition, conditions of competition can shift suddenly and asymmetrically due to their availability. Studies show that the introduction of the dynamo greatly influenced the AC/DC battle at the turn of the century, tipping the balance irreversibly towards AC.[7]

The economics of converters defies easy analysis because these products are always changing shape and their impact depends on temporary windows of opportunity. One year a converter may work only at great cost, and in the next the technology may work cheaply. One year, users invest in "anticipatory converters" to reduce the costs of a future switch between incompatible standards. In the next, a third party may enter with a new product that de facto standardizes switching. One year, a system supplier may resist the entry of all converters. In the next, de facto standards for conversion may be so well defined that a converter is longer needed. Thus, analysis tends to depend greatly on the context.

Technological innovation and industry evolution. Since so few parts of the information infrastructure have reached the stasis associated with mature product markets, standardization lies at the heart of technical change. However, because standardization may both encourage and discourage innovation in the types of products and organization of the industry, unambiguous conclusions are difficult to reach.

Does standardization encourage innovation? Because well-defined technical standards may provide component suppliers a more secure set of interfaces around which to design a product, they may encourage research and development into the design of new components for a network. Secure telecommunication transmission standards were important in hastening innovation in customer premises markets, such as facsimile machines and modems. Indeed, Noam[8] has observed more generally that the success of a communications network sponsor, such as AT&T, comes from developing and standardizing the technology of its network. Ironically, the sponsor's success lays the seeds for later third-party component competition.

On the other hand, an installed base of users may also create an unintended hindrance for innovation on a mature network. An existing substitute network may hinder the growth of a new network, because the technology embedded in much existing equipment may be inappropriate for a new application, raising its cost. In addition, minority interests may be burdened with higher costs on an existing network, but may not be large enough to justify setting up a new network. Besen[9] argued that the existing AM network hindered the post-WWII growth of the FM network.

Whether or not a network is sponsored, network capacity investment decisions determine the ultimate capability of the network. Since vendors often do not have sufficient incentives to embed interoperable technology in their equipment, one can make a case for limited government intervention aimed at guaranteeing a minimal amount of interoperability, at least to induce technical change and capacity investment. This is a frequently used argument for government regulation of electronic protocols in Internet, where fears of widespread technical chaos in the absence of minimal standardization arise.

Does standardization encourage industry concentration? Economists are equally ambivalent about the influence of standardization and technical change on a network's market structure. As noted, the factors producing less concentration are strong: network sponsors may have incentives to license their standard as a means to induce development of new components. Standards may also encourage product innovation and new entry by reducing technical uncertainty. The establishment of nonproprietary standards within the PC industry hastened the entry of multitudes of hardware, component, and software suppliers, which makes the industry incredibly dynamic and competitive today.

However, the factors leading to greater concentration are equally as strong: buyers often have strong incentives to use a single economic network. If a firm has a proprietary right over the technically superior network technology, through appropriate strategic actions (and a little luck) the sponsor can perhaps mushroom its advantages into dominant control of several technically related market niches. We can inter-

pret IBM's early success in the mainframe market with the System 360 this way. Similarly, some observers claim that Microsoft uses its control of MS-DOS and Windows for advantages in related markets (though the US Government has not quite made up its mind whether to believe these claims).

Standardization and the evolution of the information infrastructure. Economists have studied the long-run evolution of standards in a few industries rather intensively. Microprocessor markets,[10,11] computing markets,[12] VCRs,[13] and broadcasting[14] have sufficiently long and well-documented histories to point toward the following relationships between standardization and industry evolution.

First, different types of sponsorship are appropriate for different types of innovation. If a dominant firm sponsors a technology, the sponsor is more likely to innovate on a systemic level—one that influences many components at once. Typically, systemic innovations are technically complex and more easily coordinated within a single organization. RCA's shepherding of the introduction of color television (through its ownership of NBC) is one example of sponsorship working well. Sponsors of networks, however, tend to resist too much innovation, because sponsors do not want to quickly cannibalize their own products, which embody old designs. AT&T's steady, but undramatic, introduction of digital switching equipment is an often-cited example, perhaps rightly or wrongly, for both the good and bad.

In contrast, economic networks with diffuse ownership, where competitive dueling is more common, militate for greater innovation from suppliers of component parts. Component suppliers must cannibalize their own products with innovative designs just to keep ahead of the competition. One need only examine many information technology markets today to observe this trend at work. However, diffuse ownership, even combined with established producer or standards-writing groups, does not easily lead to systemic innovations, because of the difficulties of coordinating complex technical change across many organizations. Consider Unix standards development.

Second, there is a tension between the role of sponsorship in bringing about coordination and in leading to market power. When networks compete in the long run, they often become less sponsored, because many users resist the market power inherent in such sponsorship. Users choose products with wider supplier bases whenever possible, taking actions to reduce the degree of lock-in. Many users also strongly desire that at least one market institution take on a central coordination role, which leads them to a dominant firm because a single sponsor can often do a better job at coordinating a network than producer or standards-writing groups. The best example of both these tensions comes from the last 30 years of platform competition in the computing market, where users have gradually moved from sponsored

> *If a dominant firm sponsors a technology, the sponsor is more likely to innovate on a systemic level—one that influences many components at once.*

networks, such as those based on the IBM360/370 mainframe platform or the DEC VAX platform, to nonproprietary PC networks, such as those based on the Intel x86 chip and MS-DOS operating system. Intel and Microsoft have recently taken on more and more of the functions typically associated with a system sponsor, while so much of the standardization in the peripheral and software market remains nonproprietary.

In any event, prediction about the long-run evolution of an economic network is almost impossible because the success of an economic network is so closely tied to the success of the underlying technology, which is inherently uncertain (which, of course, does not prevent futuristic technologists from making predictions). Some highly touted technologies gain wide acceptance and some do not, but pinpointing the causes of success or failure is often difficult. In product markets that regularly undergo radical product innovation, it will not be clear at the outset how valuable a product or service will be, nor what the costs each technical alternative may impose on later technical developments, nor how large the network will grow as new applications develop. As a result, it is difficult to predict a market's dynamics after standardization.

For example, none of the important firms in the VCR industry in the late 1970s anticipated either the consequences for hardware competition from the development of the rental movie market, or the power of the economic links between geographically separate markets. In a more current case, technical uncertainty makes it difficult to predict whether the technical requirements implicit in ISDN (Integrated Services Digital Network) will limit or enhance competition. After all, ISDN will influence product design and network growth, which in turn may influence other factors such as tariff structures, network controls plant investment, and other regulatory decisions.

The only predictable feature of many information technology networks is that they change. It is not surprising if two snapshots of any particular market niche taken sufficiently far apart in time may reveal different firms, radically different products and applications, and even different buy-

> ## Most buyers and sellers in an evolving industry know that change will come and that its character will be unpredictable.

ers. From an individual supplier's or user's perspective, this uncertainty complicates decisions with long-run consequences, since investment in physical equipment and personnel training is expensive.

Lock-in and control of technical options. Most buyers and sellers in an evolving industry know that change will come and that its character will be unpredictable. Most product designers and users of compatibility standards thus associate potential problems with being locked-in to a narrow technical choice. One of the most interesting and least understood aspects of standardization processes is how attempts to avoid lock-in influence design decisions and market outcomes in dynamic settings.

One approach to understanding standardization activity emphasizes the value decision-makers place on having strategic flexibility—retaining a choice among many future technical options (for one line of development, see Sanchez[15]). Its starting premise is that much technology choice involves discontinuous choices among alternatives, and an important determinant of an investment is the uncertain revenue stream associated with future technical alternatives. Product designers and technology users will expend resources today so as not to foreclose technical alternatives associated with potentially large revenue streams. The greater the uncertainty at one time, the greater the value placed on keeping technical choices open over time. The value of strategic flexibility may far outweigh the value of any other determinant of technology choice.

Standards may influence a firm's decisions on whether to design a new product for a given product line, delay introducing a new product, or invest in capacity for an existing product line. A firm may choose to expend extra resources to become part of the largest possible network (by designing a standardized technical platform) because it cannot be certain which of many future designs will best suit its customers. A firm may also expend extra resources to make its products compatible with a mix-and-match network to give buyers assurance that many applications may be available in the future. A firm may hedge its bet by simultaneously employing different technical standards that permit it to reverse its commitment to a technical alternative.

Buyers will also expend resources to leave open options affected by technical uncertainties. Buyers require evidence that their technical options will remain open. The existence of many peripheral component suppliers assures buyers that an economic network caters to a variety of needs. Alternatively, users may purchase general-purpose technologies rather than an application-specific technology as a means of leaving open their options for future expansion. Elsewhere I discuss how federal mainframe computer users in the 1970s telescoped future lock-in problems into the present and made investments in "modular" programming as a result.[16]

Some, but not all, of this anticipatory activity is in society's interest. Much of it can be a nuisance and possibly wasteful. From any viewpoint, expending resources on anticipated events that do not necessarily occur is quite frustrating. This is the aspect of today's standardization processes that many firms see day to day, which may be one reason that the absence of standardization is so maligned in trade publications.

Organizational innovation or innovation by organizations

As noted earlier, there are many situations in which all component suppliers have an interest in seeing the emergence and the growth of an economic network. Yet structural impediments may produce coordination problems. The strong mutual interest all firms have in the emergence of an economic network can lead firms to forego market processes and attempt to develop standards in organizations that combine representation from many firms. How do these groups work and do they work well? Do these groups ameliorate problems identified in the short- and long-run analysis of unfettered market processes?

Consortia, mergers, and other forms of cooking by consensus. Every other day it seems that the newspaper announces the formation of another alliance or the merger of erstwhile competitors, where the participants come together to play a large role in the next phase of development of standards for some aspect of the information infrastructure. Though consortia do not have a long and well-documented history, a few examples point out some of the economic strengths and pitfalls of developing standards through these groups. Many of the same factors influence the merger of two firms, but this discussion will narrow its focus to consortia. The reader will easily see the implications for mergers. Table 3 summarizes the main features of consortia, and compares them to standards development organizations

Consortia are becoming increasingly popular in information technology industries, partially as an outgrowth of joint-research ventures. Most of these groups seem to be concerned with the future of technologies and anticipated changes in related standards. The consortium jointly operates an organization responsible for designing, upgrading, and testing a compatibility standard. The consortium lets firms legally discuss technical issues of joint interest, while osten-

Table 3. The economic role of organizations: A comparison.		
Criteria	Consortia	Standards development organizations
Motivation for formation	Strategic alliances or outgrowth of joint research venture	Professional societies
Primary benefits	May accelerate development of complementary components	Forum for discussion of issue surrounding anticipatory standards
Main hindrance to success	Strategic interests of vendors override greater interests of organization or society	Strategic interests of vendors override greater interests of organization or society
Coordination of technical change	Will coordinate change only among the subset of cooperating firms	Administrative processes tend to be slow relative to the pace of technical change

sibly avoiding anti-trust problems, but retaining considerable independence in unrelated facets of their business.

Good points to consortia. The greatest economic benefit of these groups comes from accelerated development of complementary components. Success is more likely when all the companies (who may directly compete in a particular component market) find a common interest in developing products that complement their competitive offering. Each company may offer different types of engineering expertise, whose full value cannot be realized unless combined with another firm's talents. Each firm may anticipate specializing in one part of the system that the consortium sponsors. In this respect, the incentives for firms to come together in a consortium and jointly design a standardized bundle of components resemble the incentives for several firms to independently produce a mix-and-match set of components.

However, consortia retain an interesting dynamic element. The consortium helps induce other firms to produce complementary components because the consortium's existence acts as a guarantee that a standard's future integrity will be maintained. Of course, there may still be insufficient investment in complementary products since no producer internalized the entire interest of the network, but some investment is often better than nothing, which is enough to begin development.

Problems with consortia. Consortia are not a perfect solution to coordination problems. They can easily fall prey to some of the same structural impediments that prevented network development in their absence. The experience with the development of Unix standards in the 1980s amply illustrates these weaknesses.[1] Many firms perceived strategic alliances as tools to further their own economic interests and block unfavorable outcomes. As a result, two different consortia, OSF and Unix International, originally sponsored two different Unix standards. Industry participants lined up behind one or another based on economic self-interest. More recently, different consortia (and firms) have sponsored slightly different forms of Unix, confusing the marketplace once again. While having two standards (or a few) surely is better than the multiplicity that existed before, there does not seem to be sufficient heterogeneity in user needs to merit two or more standards. Society would probably be better off with one standard, which supplier self-interest will prevent.

The other potential danger with consortia, as when any group of competing firms cooperates, is that such organizations may further the interests of existing firms, possibly to the detriment of potential entrants or users. Consortia may aid collusive activities through joint pricing decisions, or may serve as vehicles to raise entry barriers, chiefly by stifling the development of technology that accommodates development of products that compete with the products of firms inside the consortia. We will need more understanding of consortia before it is clear whether this is a common practical problem, or an unfounded fear. After all, credibly inviting development of complementary components and simultaneously deterring development of competing components may prove difficult.

Standards development organizations. One of the reasons private consortia are often unnecessary is that other well-established professional organizations serve similar functions. Many large umbrella groups that cut across many industries—CCITT, IEEE, and ASTM—have long histories of involvement in the development of technical standards.[17] These groups serve as a forum for discussion, development, and dissemination of information about standards. In the

> **Standards may raise entry barriers by stifling new entrants. The suppliers that dominate standards writing will want to further the interests of existing firms, not potential entrants or users.**

past, such groups largely codified standards determined by market processes. Today a whole alphabet soup of groups is involved with anticipating technical change in network industries and guiding their design. Their role in designing anticipatory standards takes on special urgency in economic networks in danger of locking-in to technical standards.

Do SDOs work? Standards development organizations play many useful roles in solving network coordination problems, especially those related to lack of communication. They can serve as a forum for affected parties to educate each other about the common perception of the problems they face. They can also serve as a legal means to discuss and plan the development of a network of compatible components, as well as document agreements about the technical specification of a standard and disseminate this information to interested parties. And perhaps most importantly, their standards can serve as a focal point to designers who must choose among many technical solutions when embedding a standard in a component design. These groups then are most likely to succeed when market participants mutually desire interoperability, need to establish a mechanism for communication, and need a mechanism to develop or choose one of many technical alternatives. Witness the involvement of grocers groups in the development of bar-codes for retail products.

Note that most of these organizations are "voluntary." Participating firms and individuals have discretion over the degree of their involvement. Though most firms or individuals belong to the relevant umbrella groups, their contribution of resources (and time) to development can wax or wane for a variety of technical and strategic reasons. This can lead to either extraordinary investment in the process to influence outcomes or to "free-riding" off the activities of the organization. These biases are well known, and can only be held in check by the professional ethics of the engineers who design standards.

Problems with SDOs. Voluntary standards groups are no panacea for the structural impediments to network development. They will fail to produce useful standards when the self-interest of participants prevents it in any event. Designers thus must have some economic incentives for embedding a technical standard in their product, since use is optional. A dominant firm need not follow the recommendations of a voluntary standardization group. Moreover, it is not likely to do so if it believes that it can block entry and successfully market its products without the standard. IBM's marketing of systems using EBCDIC rather than ASCII serves as one such example.[5]

Similar impasses may occur in a market with dueling technologies, although a voluntary group can play an important role in a duel. If it chooses a particular standard, it could swing the competitive balance in favor of one standard rather than another. However, each sponsoring firm may try to block the endorsement of its rival's standard as a means to prevent this result, which may effectively prevent adoption of any standard by the voluntary group. The strategies employed in such committee battles can become quite complex, ranging from full cooperation to selective compromise to stonewalling.

In addition, probably no administrative process can guide the development of a network when a slow administrative process cannot keep up with new technical developments. When events become too technically complex and fluid, a focal point easily gets lost. This problem is already arising as private telecommunications grow and private groups attempt to coordinate interconnection of their networks based on the ISDN model. One objection to ISDN is that the value from anticipating developments (on such an ambitious scale) is reduced if, as parts of the ISDN standard are written, the character of technology has changed enough to make the standard inadequate. The standard thus does not serve as a guide to component designers if the standards organization must frequently append the standard. Since no government administrative process could obviously do any better, market processes will usually predominate, coordination problems and all.

Since the decisions of voluntary groups can influence economic outcomes, any interested and organized party will make investments so as to manipulate the process to its advantage. User interests tend then to be systematically unrepresented, since users tend to be diffuse and not technically sophisticated enough to master many issues. In addition, large firms have an advantage in volunteering resources that influence the outcome, such as volunteering trained engineers who will write standards that reflect their employers' interests. Finally, insiders have the advantage in manipulating procedural rules, shopping between relevant committees, and lobbying for their long-term interests.

Committees have their own focus, momentum, and iner-

tia, which will necessarily shape the networks that arise. As a general rule, the consensus rules governing most groups tend to favor backwards-looking designs of standards using existing technology. As with consortia, standards may serve as vehicles to raise entry barriers by stifling the development of components from new entrants. The suppliers that dominate standards writing will want to further the interests of existing firms, not potential entrants or users. These biases are also well known, and are often held in check by the presence of anti-trust lawyers and, once again, the professional ethics of the engineers who design standards.

Voluntary standards organizations thus can improve outcomes for participants and society, particularly when they make up for the inadequate communication of a diffuse market structure. They provide one more avenue through which a system may develop and one more channel through which firms may communicate. They are, however, just committees, with no power to compel followers. In highly concentrated markets, their functions can be influenced by the narrow self-interest of dueling or dominant firms.

DO DECENTRALIZED MECHANISMS LEAD to appropriate standards? It is difficult to know. Neither blind faith in market processes, nor undue pessimism is warranted. Because standards can act as both a coordinator or a constraint, many outcomes are possible. Decentralized market mechanisms may produce desirable outcomes or distort them, depending on the market structure, chance historical events, and changes in the costs of technical alternatives. Diffuse market structures produce coordination problems and communication difficulties, but also much innovation. More concentrated market structures will alleviate some of the communication problems, but strategic interests will distort incentives away from optimal outcomes. Administrative processes may ameliorate communication problems, but internal political battles will distort outcomes in other ways.

In my view, the progressive decentralization of decision-making in information technologies away from a few sponsors, such as AT&T and IBM, has to be good in the long run. This decentralization has unleashed an unmanageable variety of entrepreneurial activity. There is a natural (and some-

The problems associated with standardization are an unfortunate, but bearable and necessary, cost associated with such change.

times legitimate) desire to want to manage and slow down the massive changes that accompany such entrepreneurial activity. However, such desires should not dictate the pace of change. Dynamism leads to economic growth and development and fantastic technical possibilities. The problems associated with standardization are an unfortunate, but bearable and necessary, cost associated with such change.

If the dynamism of the last few decades is any guide to the future, no one should lament the much-maligned present state of affairs too loudly. Recent history makes me cautiously optimistic about the role of decentralized market mechanisms in guiding standardization development within economic networks today, provided that it is properly modified by a professional standard setting process. I look forward to analyzing future developments; they will be as important as they are interesting. ⬙

References

1. G. Saloner, "The Economics of Computer Interface Standardization: The Case of Unix," *Economics of Innovation and New Technology*, Vol. 1, Harwood Academic Publishers, New York, 1989, pp. 135-56.
2. F.M. Fisher, J.J. McGowan, and J.E. Greenwood, *Folded, Spindled, and Mutilated: Economic Analysis and US vs. IBM*, MIT Press, Cambridge, Mass., 1983.
3. J. Rotemberg and G. Saloner, "Interfirm Cooperation and Collaboration," *Information Technology and Organizational Transformation*, M.S. Morton, ed., Oxford University Press, New York, 1991, pp. 95-121.
4. S.M. Besen and G. Saloner, "Compatibility Standards and the Market for Telecommunications Services, " *Changing the Rules: Technological Change, International Competition and Regulation in Telecommunications*, R.W. Crandall and K. Flamm, eds., The Brookings Institution, Washington, D.C., 1988, pp. 177-220.
5. G. Brock, "Competition, Standards, and Self-Regulation in the Computer Industry," in *Regulating the Product: Quality and Variety*, R. Caves and M. Roberts, eds., Ballinger Publishing Co., Cambridge, Mass., 1976.
6. G. Brock, "Dominant Firm Response to Competitive Challenge: Peripheral Equipment Manufacturers' Suits Against IBM (1979-1983)," in *The Anti-trust Revolution,* J.E. Kwoka and L.J. White, eds., Scott, Foresman, and Co., Glenview, Ill. 1989, pp. 160-182.
7. P.A. David and J.A. Bunn, "The Economics of Gateway Technologies and Network Evolution: Lessons from Electrical Supply History," *Information Economics and Policy*, No. 3, 1988, pp. 165-202.
8. E.M. Noam, "Network Tipping and the Tragedy of the Common Network: Theory for the Formation and Breakdown of Public Telecommunications Networks," Columbia Inst. for Tele-Information, Columbia Business School, Columbia Univ., New York, Feb. 1991.
9. S.M. Besen, "AM vs. FM: The Battle of the Bands," *Industrial and Corporate Change*, Vol. 1, No. 2, Oxford Univ. Press, 1992, pp. 375-396..
10. J. Wade, "Dynamics of Organizational Communities and Technological Change: An Empirical Investigation of Technological Variation, Innovation, and Performance in the Microprocessor Market," doctoral dissertation, Walter A. Haas School of Business, University of California, Berkeley, Calif., 1993.
11. G.M.P Swann, "Product Competition," Microprocessors, *J. Industrial Economics*, Vol. 34, No. 1, Basil Blackwell, Oxford, England, Sept. 1985, pp. 33-53.
12. T. Bresnahan and S. Greenstein, "Technological Competition and the Structure of the Computer Industry," CEPR Discussion Paper No. 315, Stanford Univ., Stanford, Calif., June 1992.
13. M.A. Cusumano, Y. Mylondadis, and R. Rosenbloom, "Strategic Maneuvering and Mass Market Dynamics : The Triumph of VHS over BETA," Working Paper 91-048, Harvard Business School, Cambridge, Mass., 1991.
14. B. Owen and S. Wildman, *Video Economics*, Harvard Univ. Press, Cambridge, Mass., 1992.
15. R. Sanchez, *Strategic Flexibility, Real Options, and Production Strategy*, doctoral dissertation, Dept. of Economics, MIT, Cambridge, Mass., June 1991.
16. S. Greenstein, "Lock-in and the Costs of Switching Mainframe Computer Vendors: What Do Buyers See?," University of Illinois, Faculty Working Paper 91-0133, Political Economy Series #48, Urbana-Champaign, Ill., 1991.
17. D. Hemenway, *Industrywide Voluntary Product Standards*, Ballinger, Cambridge, Mass., 1975.
18. S. Greenstein, "Invisible Hands and Visible Advisors: An Economic Interpretation of Standardization," *J. Am. Soc. Information Science*, Vol. 43, No. 8, John Wiley & Sons, New York, Sept. 1992, pp. 538-549.

Further Readings

B. Arthur, *Competing Technologies: An Overview, in Technical Change and Economic Theory*, Dosi, et al., eds., Pinter Publishers, London, 1988, pp. 590-607.

Y. Baba and K. Imai, "Systemic Innovation and Cross-Border Networks: The Case of the Evolution of the VCR Systems, in Entrepreneurship, Technological Innovation, and Economic Growth: Studies in the Schumpeterian Tradition," F.M. Scherer and M. Perlman, eds., Univ. of Michigan Press, Ann Arbor, Mich., 1992.

S.M. Besen, "The European Telecommunications Standards Institute," *Telecommunications Policy*, Vol. 14, Dec. 1990, pp. 521-30.

S.M. Besen and L.L. Johnson, *Compatibility Standards, Competition, and Innovation in the Broadcasting Industry*, Rand Corp., Santa Monica, Calif., 1986.

T. Bresnahan and A. Chopra, "Users' Role in Standard Setting: the Local Area Network Industry," *Economics of Innovation and New Technology*, Vol. 1, 1989, pp. 97-110.

G. Brock, "The Regulatory Change in Telecommunications: The Dissolution of AT&T," in *Regulatory Reform: What Actually Happened?*, L.W. Weiss and M.W. Klass, eds., Little, Brown and Co., Boston, 1986.

C.F. Cargill, *Information Technology Standardization: Theory, Process, and Organizations*, Digital Press, Bedford, Mass., 1989.

P.A. David, "Some New Standards for the Economics of Technological Standardization," in *The Economic Theory of Technological*

Policy, P. Dasgupta and P.L. Stoneman, eds., Cambridge Univ. Press, London, 1987.

P.A. David and S. Greenstein, "The Economics of Compatibility Standards: An Introduction to Recent Research," *Economics of Innovation and New Technology*, Vol. 1, Harwood Academic Publishers, New York, 1990, pp. 3-41.

P.A. David and W.E. Steinmueller, "The ISDN Bandwagon is Coming—Who Will Be There to Climb Aboard?: Quandaries in the Economics of Data Communication Networks," *Economics of Innovation and New Technology*, Vol. 1, Harwood Academic Publishers, New York, 1990, pp. 43-62.

P.A. David and W.E. Steinmueller, "The Economics of Compatibility Standards and Competition," TOP Working Paper, Stanford Univ. Dept. of Economics, Stanford Calif., Jan. 1992.

J. Farrell and G. Garth, "Competition, Compatibility, and Standards: The Economics of Horses, Penguins and Lemmings," in *Product Standardization as a Tool of Competitive Strategy*, H. Landis Gabel, ed., North-Holland, New York, 1986.

J. Farrell, "Competition With Lock-in", No. 8722, U.C. Berkeley, Economics Dept., Jan. 1987.

"Information Technology Standards: The Economic Dimension, Committee for Information, Computers and Communication Policy," OECD, Paris, 1991.

B. Kahin, *Building Information Infrastructure*, McGraw-Hill, New York, 1992.

C.P. Kindleberger, "Standards as Public, Collective and Private Goods," *Kyklos*, Vol. 36, Verlag, Basil, Switzerland, 1983, pp. 377-397.

R.N. Langlois and P.L. Robertson, "Innovation in Modular system: Lessons from the Microcomputer and Stereo Component Industries," mimeo, Univ. of Connecticut Dept, of Economics, Storrs, Conn., Apr. 1990.

W. Lehr, "Standardization: Understanding the Process," *J. Am Soc. Information Science*, John Wiley & Sons, New York, Sept., 1992.

A. Link and G. Tassey, "The Impact of Standards on Technology-Based Industries: The Case of Numerically Controlled Machine Tools in Automated Batch Manufacturing," in *Product Standardization and Competitive Strategy*, H.L. Gabel, ed., North Holland, Amsterdam, 1987.

R. Noll and B. Owen, *The Anti-Competitive Uses of Regulation: United States v. AT&T (1982)*, J. E. Kwoka and L. J. White (eds.), Scott, Foresman, and Co., Glenview, Ill., 1986, pp. 290-337.

S. Postrel, "Competing Networks and Proprietary Standards: The Case of Quadraphonic Sound," *J. Industrial Economics*, Vol. 39, No. 2, Dec. 1990.

P.L. Robertson and R.N. Langlois, "Modularity, Innovation, and the Firm: The Case of Audio Components, in Entrepreneurship, Technological Innovation, and Economic Growth: Studies in the Schumpeterian Tradition," F.M. Scherer and M. Perlman, eds., University of Michigan Press, Ann Arbor, Mich., 1992.

M.B. Spring, "Information Technology Standards," in *Annual Review of Information Science and Technology*, M.E. Williams, ed., Vol. 26, Learned Information for ASIS, Medford, N.J., 1991.
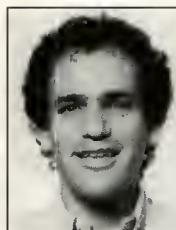
Shane Greenstein is assistant professor of economics at the University of Illinois at Urbana-Champaign, where he teaches courses in micro-economics, industrial economics, regulation, and the economics of technical change. He has written about competition on the local telephone loop, the economics of user and vendor lock-in, competition between alternative computing platforms, economic methods for evaluating improvements in computing, and the economics of federal procurement of computer systems.

Greenstein received his BA from the University of California at Berkeley and his PhD from Stanford University. He is a member of the American Economics Association and a faculty research fellow with the productivity group at the National Bureau of Economic Research.

Address questions concerning this article to Shane Greenstein, Department of Economics, University of Illinois at Urbana-Champaign, 1206 S. Sixth Street, Box 162, Champaign, IL 61820; grnstn@vmd.cso.uiuc.edu.

Reader Interest Survey

Indicate your interest in this article by circling the appropriate number on the Reader Service Card

Low 162   Medium 163   High 164

Forming, Funding, and Operating Standard-Setting Consortia

In general, the greater the proprietary, or strategic, motivations for the formation of a consortium, the less likely the consortium will succeed and sustain itself. In contrast, nonstrategic consortia continue to provide useful de facto standards to the industry at large. Both types of consortia can benefit from an understanding of how their structures, governance, funding, and other attributes contribute to their success or failure.

Andrew Updegrove

Lucash, Gesmer & Updegrove

Computer companies formed so many industry associations in 1988 that several commentators dubbed it the Year of the Consortium. Few of these organizations intended to develop true de jure standards. Instead, most were marketing partnerships formed in response to intense competition among the largest companies, each jockeying to dominate or at least maintain its position in volatile market segments such as Unix platform sales. Despite their overtly commercial purposes, these organizations frequently found the development of de facto standards and specifications crucial to achieving their ends.

Although the formation of these de facto consortia has recently slowed down, some of the earlier ones live on and continue to maintain their specifications and standards. Often these standards incorporate de jure standards. For example, the X Window System standard incorporates certain IEEE standards. In some cases, de jure standards bodies accept de facto standards or link them to their own standards—for example, Object Management Group's widely recognized Common Object Request Broker (CORBA) specification. Some consortia, such as X/Open, develop software environments using a mix of de jure and widely supported de facto standards.

Thus, de facto standard-setting organizations play an important role in relation to de jure standard-setting organizations such as ASC X3 (an ANSI subcommittee) and the IEEE. On the one hand, if a de jure organization is satisfied that a de facto organization is providing a useful standard, it may opt as a practical matter not to develop a competing standard. As a result, a de facto standard-setting organization effectively acquires control of an important area of technology. On the other hand, a de facto body's endorsement or incorporation of a de jure standard augments that standard's effectiveness.

An examination of these important effects of de facto consortia can provide insights into today's high-stakes, high-technology wars. This article looks at what these groups set out to achieve and how successful they were. In addition, it explains why and how various types of consortia were formed and outlines the financial, structural, and legal factors involved in setting up and operating a successful consortium.

Why form a consortium?

To some extent, standard-setting consortia are new wine in old bottles. Trade associations have existed for many years to promote the interests of their members. (For their history, see Garcia, this issue, p. 28.) In the building trades, for example, associations traditionally engaged in lobbying and other activities primarily intended to protect the members' business and jointly market their wares. In addition, association activities often involved influencing the development and

promotion of building standards for local adoption, again primarily for the benefit of their own members. In the era of high technology, this traditional role has not been abandoned. For example, a new trade association to promote the X Window System technology is forming independently of the X Consortium, which MIT has just spun off to continue developing the X Window System standard.

Of greater interest than traditional trade associations are certain new types of consortia, which focus primarily on the development of standards or new technology, with varying degrees of proprietary importance to their own members.

Research consortia. A number of high-profile consortia, such as Sematech and MCC (Microelectronics and Computer Technology Corporation), emerged largely as a result of heightened national procompetitive concerns. A short-lived research consortium called US Memories, for example, was established to permit semiconductor vendors to cooperate in producing low-cost DRAM chips. Some of these groups succeeded, some failed to coalesce, and some wandered from their original purposes as they sought to sustain themselves. In some cases the problems (loss of dominance in DRAM chip production, for example) that helped launch the enterprises abated. Then smaller groups of companies (sometimes including Japanese or European partners), without the high-profile, highly political trappings of their predecessors, entered into joint ventures to produce specific products.

Specification consortia. Groups such as the VXIbus Consortium (VXI) and the MIDI Manufacturers Association (MMA) are primarily concerned with providing a usable, robust standard for the benefit of an entire industry. Essentially apolitical, they direct most of their efforts to developing and supporting a specification using member (not paid) staff, and they consequently have small budgets. At their best, they successfully avoid proprietary influences and implement the best technology to produce sensible, effective, practical standards.

Vendors established and funded many specification consortia. End users established others, such as the CAD Framework Initiative, to lower acquisition costs of products that would otherwise be based on proprietary, incompatible technologies. Such groups often attempt to fill technical gaps in important niche industries too small to merit the attention of recognized standard-setting bodies such as the IEEE. For example, the MMA's goal was to provide an essential interface standard enabling the electronic music industry to develop in an accelerated and orderly fashion.

Strategic consortia. As used here, the term *strategic consortium* refers to a group initially formed and funded by a limited number of companies to promote the adoption of a particular technology (such as a chip architecture) as an "open" technology. Often, the primary mechanism for achieving this end is the development and promotion of a standard. The efficacy of this mechanism in helping to establish the under-

Some of these groups succeeded, some failed to coalesce, and some wandered from their original purposes as they sought to sustain themselves.

lying technology has varied widely in practice.

Perhaps the greatest and most interesting flowering of strategic consortia occurred in the Unix marketplace. There we find consortia devoted to the operating system itself, such as Unix International and OSF (Open Systems Foundation). In addition, we find consortia concerned with chip architectures, including 88open, ACE (Advanced Computing Environment), and Sparc International, and with software environments (X Consortium). These consortia have one or more of the following characteristics and objectives:

- They were usually formed by hardware vendors or chip manufacturers seeking a market share in either of two ways: 1) by controlling (or preventing a competitor from controlling) the evolution of Unix or 2) by fostering the rapid porting of massive amounts of software to a specific chip environment to encourage adoption of that chip by as many manufacturers as possible.
- They employed "hard" marketing tactics such as the prevention of the adoption of a proprietary standard. A prominent example of such efforts is the virtually overnight industry adoption of the MIT X Window System standard to parry the spread of Sun Microsystems' NeWS technology. Another example is the formation of the massively funded OSF to counter the benefit that some companies feared Sun might reap from AT&T's ownership of Unix, following AT&T's purchase of a significant block of Sun stock (later sold). The latter example demonstrates the varied life cycle of some consortia. AT&T eventually conveyed Unix to Unix Systems Labs and formed another consortium, Unix International, to control and promote the operating system more publicly. More recently, Novell acquired Unix Systems Labs and has made the "Unix" trademark widely available. Now, OSF—with a budget in the tens of millions of dollars and license fee income probably covering less than half that amount—is engaged in a process of self-examination to determine what its long-term mission should be.
- They used "soft" marketing techniques such as claiming
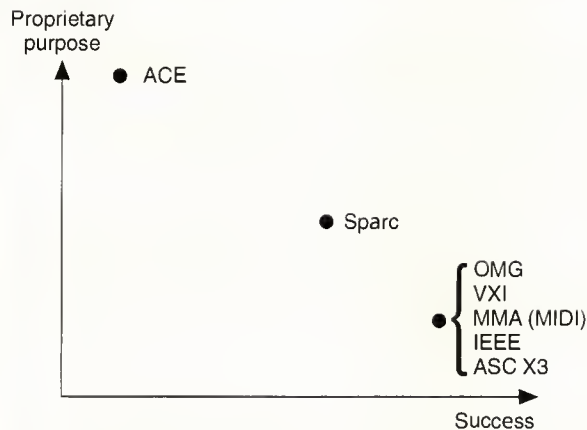
Figure 1. Correlation of consortium's level of proprietary purpose and successful development of standard or specification.

Figure 2. Correlation of standard's nonproprietary importance to industry and consortium's success.

(accurately or otherwise) the development of standards, the existence of large amounts of software embodying those standards, and the achievement of real interoperability and open systems status.

In some cases, consortia had roots of one type and grew branches of another. The X Consortium is an example. The X Windows technology had been under development at MIT since 1984, and the industry began adopting it as a standard in 1987. In a sense, the X Consortium began as a strategic consortium, formally established in 1988 out of concern over a Sun Microsystems software product initiative. The organization then enjoyed over five years of relatively ecumenical existence. That is, it evolved into a nonstrategic specification consortium dedicated to the evolution of a usable standard. Today, as the X Consortium leaves its home at MIT to become an independent entity, Microsoft's release of Windows NT gives the organization new significance as a strategic vehicle for maintaining the vitality of Unix and the X Window System standard.

Similarly, OMG (Object Management Group) was founded to develop a family of standards to facilitate the adoption of object-oriented programming methods and products. Today, under its executive director Chris Stone, the organization has broadened its scope to a wide range of activities dedicated to nurturing and promoting the fledgling industry. These activities include annual trade shows in six countries and four continents, training sessions and publishing ventures. and innovative joint marketing and electronic distribution projects. At the same time, OMG has become recognized as the preferred source of many types of object-oriented programming specifications and therefore an industry focal point.
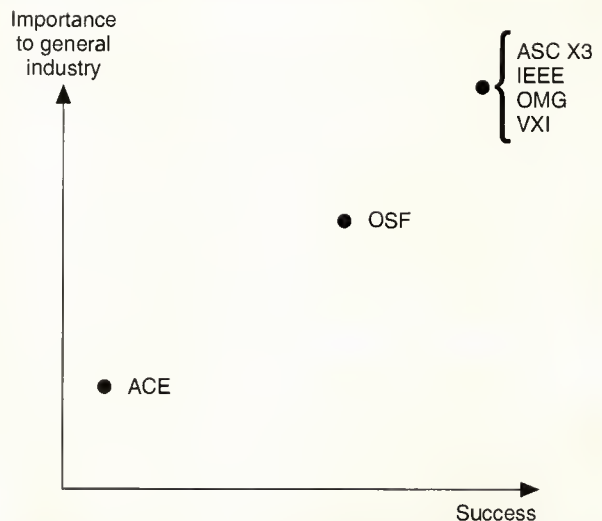
For several consortia, the key to success was that they were "ahead of the curve." OMG, for example, set out to establish a standard before any significant company obtained a vested interest in promoting its proprietary technology. As a result, the industry could follow OMG's lead and cooperate to achieve the best technical result for mutual benefit.

Unfortunately, the reverse has also been true. The first RISC chip consortium, 88open, was brilliantly successful on the technical level, achieving the highest degree of true interoperability of any of the competing organizations. But its technical success was insufficient to overcome certain other handicaps. For example, the chip technology upon which 88open was based was introduced in the marketplace too late to achieve a momentum of adoption from which the group might have benefited. The new PowerOpen Consortium (styled in many ways on the 88open Consortium) is virtually assured of avoiding this fate; both IBM and Apple (among others) have already announced the use of Motorola's PowerPC series RISC chips in their products. The sales momentum those companies represent should be more than sufficient to ensure speedy and efficient porting of massive amounts of software, even if the development of standards and test suites is not as effectively achieved.

A consortium's success tends to correlate strongly and inversely to the degree of proprietary advantage that its founders sought to gain (see Figure 1). This phenomenon results from a variety of factors. For one thing, there are always more losers than winners in industries dominated by a small number of large players. Second, the word *standard*, stretched to the limit of anyone's most liberal commercial definition (such

as VCR formats or PC operating systems), cannot usually accommodate more than a few instantiations on a long-term basis. Moreover, many strategic consortia were formed as much for publicity as for definitive purposes, were announced before a clear plan of action had evolved, or fell apart on the shifting strategic sands upon which they were based. (An example is the ACE consortium, which suffered from other problems as well.) Finally, strategic consortia are more likely to be formed by companies that perceive that they are already at a disadvantage, and the odds are therefore against success from the beginning.

Similarly, a consortium's success correlates highly and directly to the absence of a perceived proprietary advantage to any individual company. OMG, for example, was formed before more than a handful of products were even in the design stage. Other successful consortia exploited a high degree of agreement among a critical mass of companies as to the existence of a common enemy. Such was the case in the X Consortium's response to Sun. Not surprisingly, the most successful and stable consortia are those whose purposes are most beneficial to the industry as a whole (see Figure 2).

How to form a consortium

To succeed, a consortium must carefully address a number of issues. How it handles these issues depends on the goals, competitive positions, and other unique features of the member companies.

Funding. Frequently, the first reality the consortium must confront is money. When the group's goals are ambitious—including certification as well as standards development, for example—funding needs can be extreme. (See Figure 3; brackets indicate ranges of funding required by groups of each type.) These needs can be met only by enrolling very large memberships (X Consortium, OMG) or requiring very large contributions from individual members (OSF and Power Open). Typically, large-budget consortia are strategic consortia, with proprietary vendors (computer or chip manufacturers) contributing the lion's share of funds (see Figure 4). When funding is of this magnitude, an organizational structure ensuring that funds contributed are tax deductible and not regarded as capital contributions is especially important.

A second funding concern is to ensure that all companies whose participation is essential to success become members. Accordingly, a consortium's dues structure commonly reflects this need, with some classes of members bearing a disproportionate contribution obligation. These members are motivated to take up this burden by the expectation of reaping a disproportionately high economic benefit in the marketplace. For example, sponsor-level membership in PowerOpen (whose main goal is to foster rapid porting of software to the PowerPC environment) requires $250,000 in annual dues and initiation dues of $750,000. Independent software vendors (ISVs) however may participate at a limited level for only $100 a year.
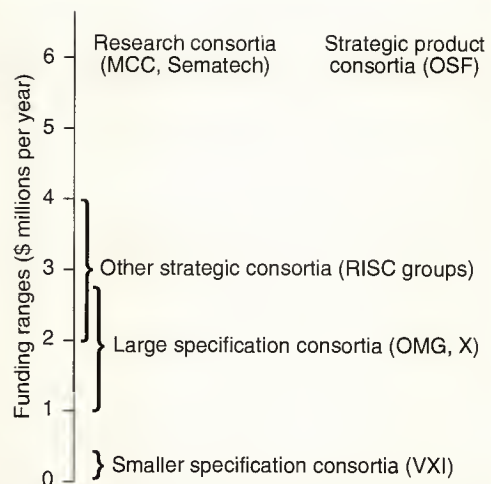
Figure 3. Approximate funding needs of various types of consortia. Funding levels of research and strategic product consortia are at least $6 million and rising rapidly.
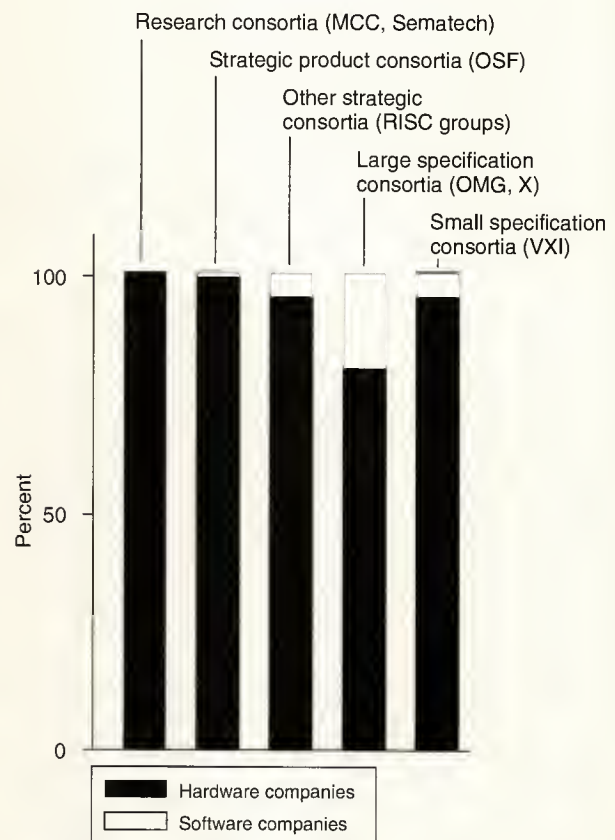
Figure 4. Sources of consortium funds.

Table 1. Typical consortium dues structures.

	Member company's annual revenue ($M)/dues ($K)			
Type	<50	50-500	>50	>500
Strategic consortium				
Type of member				
Sponsor	750 regardless of revenues			
Principal	50	150	—	350
Associate	10	15	—	25
End user	10	10	—	10
ISV	2	2	—	2
Total budget	$4M/yr (3 sponsors, 10 principals, balance of income from other members, publications)			

	Member company's annual revenue ($M)/dues ($K)			
Type	<50	50-500	>50	>500
Strategic consortium				
Type of member				
Full	25	—	50	—
Associate	5	—	10	—
End user	5	—	5	—
Individual, university	0.1	—	0.1	—
Total budget	$2.5M/yr (50 full members, 150 associates, balance of income from other members, publications, meeting fees, cooperative advertising)			

mains the focus of the dues-paying members, the consortia pay little attention to whether their collective investment merits the development of any long-term goals as well. Often, in fact, it does not, and the consortium should promptly and efficiently disband when it has achieved the grand design or failed.

In other cases, however, such organizations could have useful long-term lives as user groups, providers of certification services, and the like. Unfortunately, when the major objective has been realized or abandoned—or marketplace or technology shifts have made the objective irrelevant—it is often too late to downsize the organization to a sustainable level. At that point, the original strategic proponents are no longer interested in providing the funds necessary to pursue humbler objectives. Organizations that successfully make such a transition are those that have used their high-profile years to quietly backfill their coffers through establishing long-term activities such as publishing and joint marketing programs.

The most stable and longest-lived consortia require modest annual fees (on the order of $5,000 for companies with annual sales under $10 million, $25,000 for companies with sales from $10 million to $500 million, and $50,000 for companies with sales over $500 million). Members of these consortia tend not to send upper-level management (being only human) to meetings or to examine membership benefits as aggressively or frequently as members of consortia with higher dues. Consequently, the consortium management can steer the organization more independently—usually toward a long-term, expanding organization.

Where this process works best, innovative executive directors regularly roll out new programs to keep members happy and renewing. Where the process doesn't work, the organization rightfully withers and dies.

Technological development. Consortia use numerous methods for developing a technology, each with its own advantages. The spectrum includes

- total in-house development (88open designed and built almost all its test suites),
- the use of contract parties (such as UniSoft or ApTest, an 88open spin-off company),
- the use of contributed employees (a method used by CAD Framework Initiative), and
- a request-for-technology (RFT) process (the OSF and OMG method).

Indeed, all types of consortia are likely to maximize revenue by offering various levels of membership (as discussed in greater detail later under "Governance"), with higher dues required of membership levels that receive early access to technology and other benefits. Many consortia also charge different rates for companies with different revenue levels, to permit smaller companies to participate (or to charge their larger members more, depending on your point of view). Table 1 shows the typical dues structures of strategic and specification consortia.

Large-budget consortia suffer from a disease peculiar to the species that might be called the cliff syndrome. Simply put, the larger the budget, the greater the likelihood that the organization will hurtle at full-budget speed off a cliff of its own making and into oblivion. This syndrome is particularly prevalent among strategic consortia. The origin of the problem is that such consortia are formed to achieve a specific grand objective, at great expense. While this objective re-

Some consortia, such as the X Consortium, use a mix of sources, including internal staff, members working outside the consortium, and technology contributions from members. Table 2 lists methods of technology development used by several consortia.

A number of advantages and disadvantages stand out when we compare these methods. In-house development maximizes control and planning and minimizes time to completion, but it requires the highest levels of funding. Contributed employees are sometimes hard to motivate. They may be preoccupied with what is going on at their home office (if they are on site at the consortium), and often their employers don't give them ample time or credit for properly completing their assigned consortium tasks. Issuing an RFT can reap a rich offering of responses, but working through them to a final acceptance can be a laborious process. It also requires the most scrupulous legal scrutiny to avoid antitrust problems and often invites spirited discussions among members. In one instance, our firm (as legal counsel to a consortium) had to recommend that we open sealed ballots on a final technical decision in secret and announce only the final tally. We took this precaution to free smaller members from concerns over retribution from more powerful vendors, who had broadly hinted that a vote against their submission could have adverse business consequences.

Governance. When our firm first began representing consortia, we examined a wide variety of organizational forms from various jurisdictions. We eventually settled on the Delaware not-for-profit, nonstock membership corporation. This structure permits members to join and leave with ease, provides members (and their counsel) with a convenient and familiar body of corporate law, and provides a high degree of operational flexibility. Most of the largest East Coast-based consortia are now formed on this model, while many West Coast consortia use a similar format under California law. The structure has stood up well in practice.

The by-laws and charter are the heart and soul of a consortium. Although a few important rules may be stated in the membership application, most of the regulations and rights of the organization appear in the by-laws and charter. To a large extent, these legal documents (along with the rules of technical committees) will determine eventual success or failure. They determine whether the organization is easily managed, whether it avoids needless exposure under the antitrust laws, whether its members feel fairly represented and therefore renew their memberships, and whether it is sufficiently flexible to evolve and flourish.

For example, a crucial factor for a stable, sustainable organization is its method of representing members on the board of directors. Organizations we have represented have developed a variety of effective formulas (some organizations have changed from one method to another at several points in their evolution). These include economic models (those who pay the highest dues get the board seats) and arbitrary solutions (the first members to join get the seats, while later members stand in line for an opening). Other models place a premium on democratic values or objectivity. For example, a certain number of seats may be allocated to each type of member—ISVs, hardware manufacturers, and academics—to ensure that all interest groups are heard from. Conversely, only nonmembers may be allowed to become directors so that standards adopted are "pure."

Most (but not all) consortia have different membership categories, each category having different rights. These rights may include voting for or nominating directors, participating in committees, early access to technology or standards, and reduced-price or free access to standards, certification, or other services. In some cases, members must pay higher dues to exercise these rights.

To succeed, a consortium must set dues appropriate to its benefits, and organizations sometimes go through some trial and error before arriving at the right formula. Later the consortium may need to make periodic readjustments to maintain member satisfaction; many consortia have gone through periods of declining membership before realizing that a recalibration of dues and services was necessary. Table 3 (next page) summarizes some of the more important rights of selected membership classes in several consortia.

Typically, consortium activities are performed by technical committees, work groups, and special-interest groups, as well as by nontechnical governing bodies, such as business, audit, and executive committees.

US tax issues. Many consortia elect to operate as tax-exempt trade associations under Section 501(c)(6) of the Internal Revenue Code. A few also qualify as public charities under Section 501(C)(3), which offers more advantages, but more restrictions (both federal and state) as well. As such, a consortium may carry out most of its activities without payment of income taxes on dues or related income. However, the

Table 2. Sources of technology (standards, specifications, test suites).					
Organization	In-house	Member contrib.	General RFT	Contract	Contributed employee
88open	X			X	
OMG		X	X		
X Consortium	X	X			X
Sparc Int'l	X	X		X	

Table 3. Selected membership-class rights.					
Right	88open	PowerOpen	X Consortium	OMG	Sparc Int'l
Automatic board seat (for highest fee level)	Yes	Yes	No	No	Yes
General voting (number of classes eligible/total number of classes)	2/4	2/3	2/3	1/5	3/5
Automatic technical committee seat or shared seats (number of classes eligible/total number of classes)	2/4	2/3	3/3	2/5	3/5
Early access (in some form) to technology	No	Yes	Yes	Yes	Yes

IRS holds that some activities, such as the inclusion of paid advertising in consortium publications, stand outside the organization's tax-exempt purposes and that income from these activities is "unrelated business taxable income." Taking in too much of this taxable income may endanger the consortium's overall tax-exempt status. Accordingly, if these activities are important to the consortium, it is often well-advised to carry them out through a separate corporation. Fortunately, the IRS usually regards many common fee-generating activities, such as certification testing and standards publishing, as related to the consortium's purpose and thus falling within tax-exempt parameters.

Although tax avoidance (as compared to tax evasion) is legal, and some would consider it a hallowed part of the American psyche, consortium founders should keep in mind that tax avoidance is not their primary purpose. When a consortium has an admittedly strong proprietary purpose, obtaining tax-exempt status is usually impossible without crippling that very purpose. Similarly, when an organization intends to embark upon many taxable activities, the tax structuring (such as creating multiple subsidiaries) necessary to avoid taxation of other activities can rise to such a level of artificiality and inconvenience that it ultimately outweighs the benefit of the tax exemption.

On balance, for a true standard-setting organization, seeking tax-exempt status is usually worthwhile. A group with a proprietary purpose or complex plans might better forgo filing or abandon tax-exempt status later, if it can do so without adverse consequences. Organizations without tax-exempt status can minimize or eliminate taxes by operating as closely as possible to a break-even point.

Where feasible, consortia often seek multiyear commitments from members to facilitate long-term planning on issues such as hiring of management and entering into leases. OSF's multimillion-dollar, multiyear commitments—recently renewed by many original members—continue to lead the pack in this regard. Obviously, long-term issues are most troublesome for the "cliff syndrome" consortia mentioned earlier. Long-term planning is easiest for stable consortia with large member bases and smaller annual dues, such as OMG, now the world's largest software consortium with over 320 members. Table 4 lists the tax status of several consortia.

Antitrust issues. By definition, consortia are combinations of competitors. As such, they must operate with careful regard to US federal and state antitrust laws. Individuals and corporations found to have violated these laws face criminal liability as well as the possibility of treble-damage suits from private parties. As this area of law is complex, the following is a brief and superficial review of only a few relevant antitrust issues.

Most consortia can safely operate by being mindful of several rules that are, on reflection, logical and obvious. They include not operating in a manner that might shut nonmembers out of the marketplace, not excluding eligible companies from membership, and not trading price information or seeking to set or influence prices. Sometimes, however, a consortium wants to carry out a new practice for which there are no (or only poorly analogous) precedents. In such cases, the consortium must undertake a careful analysis to ensure that the practice incurs no antitrust risks. For example, the

consortium may wish to charge nonmembers high prices for test suites created at great expense by the consortium. If these prices are not a barrier to entry into the marketplace, bear a logical relationship to the development costs of the test suites, and do not compel membership in the consortium (by costing the same, or more, than membership dues), they will usually be upheld. Generally, companies can join consortia without fear of adverse judgments if the consortia are prudently managed and are monitored by legal counsel.

Happily, the law presumptively deems standard setting an appropriate joint activity. Engaging in certification services is less explicitly favored, but is permitted so long as proper controls are established to ensure nondiscriminatory pricing and access to testing. Generally, consortia may charge nonmembers higher fees than members for certification testing and other services and products, provided dues already paid by members were spent to establish the services or products in question.

One way a consortium can minimize risk under the antitrust laws is to register under the National Cooperative Research and Production Act (NCRPA) of 1993. The act protects organizations engaged in certain types of activities (such as, arguably, standards development) from the risk of treble damages and liability for a plaintiff's attorney fees. In June 1993, Congress expanded the activities covered by the act. Although the act still does not cover all types of activities a consortium is likely to engage in, it does cover a large enough number of typical activities to make registration worthwhile for many consortia.

However, the NCRPA requires public disclosure of members. Disclosure may not be acceptable for a strategic consortium whose members have not yet announced their commitment to a particular proprietary technology. For these members, the wish to prevent their competitors from discovering their strategic direction may outweigh the benefits of NCRPA registration. Moreover, a number of inartfully conceived or politically influenced aspects of the act make it difficult to reliably predict whether it will protect against liability in a given situation or for a given member. For example, whether some non-US members of a consortium can take advantage of some of the act's protections is probably impossible to determine.

Finally, there is no guarantee that the predominantly laissez-faire antitrust policies of the previous two administrations will continue (in fact, the Clinton administration has already indicated that we can expect at least some stiffening of enforcement). Even under the Bush administration, OSF and MicroSoft were the subjects of federal antitrust investigations. Rumors implied that the government also undertook an investigation of high-technology consortia in general (thus far without public result). Although procompetition activities presumably will continue to receive at least some measure of government sympathy, if not outright legislative encourage-

Table 4. Regulatory exemptions.		
Organization	Tax exempt	NCRPA-registered
88open	No	No
OMG	Yes	No
X Consortium	Yes	Yes
Sparc Int'l	No	No

ment, consortia should not forget the potential activities of private litigants (such as competitors). For example, OSF was sued by private parties concurrently with the commencement of the government investigation, and it is likely that the government inquiry was inspired by private party complaints.

Thus, every consortium should have an active antitrust oversight program, including the distribution of appropriate educational materials to officers, directors, and members. Where the budget permits, legal counsel should attend directors' meetings. Potentially discriminatory activities such as the adoption of technology should be the subject of written policies that are reviewed by counsel and faithfully implemented.

Technology ownership. Consortia may or may not plan to develop technology when they are formed, but nevertheless they often create intellectual property as they evolve. Since members may come and go, and the organization itself may at some point dissolve, consortia must carefully consider the current and eventual ownership of the technology they develop.

The ownership problem also arises when consortia make use of contributed employees (a frequent practice). Usually, those employees are bound by agreements that all results of their work will belong to their employers. Similarly, many consortia accept contributed technology, which, as part of a specification, a standard, or an implementation of a standard, becomes freely available to the world. When technology is developed over time through the contributions of many individuals and corporate members, unscrambling the ownership of that technology and the ultimate liability for any error or infringement can become a nightmare.

The history of the X Window System standard exemplifies how complicated such matters can be. The standard was initially developed at MIT by MIT staff. Later, industry supported the standard with financial contributions to the X Consortium. Member and nonmember technical contributions were often accepted, modified, and incorporated into the standard although actual ownership of the technology contributed was not necessarily transferred. At the same time,

the MIT staff was developing further technology within the X Consortium.

The consortium licensed certain software representing the standard to the world at large at a minimal fee, thus releasing actual code to commercial users, who incorporated it in their products. The software distributed by the consortium also included third-party commercial software (such as fonts), which was similarly distributed and reused. To complete this complex picture, MIT is now transferring its ownership rights to the new, independent X Consortium.

When a consortium's espoused goal is to develop and support an open, nonproprietary standard, it must avoid incurring any clouds on ownership and control, even under convoluted developmental conditions like those just described. Similarly, it will wish to avoid incurring any liability to members or third parties who adopt consortia standards or software. Therefore, every consortium should institute a careful program of technology procurement, documentation, protection, labeling, and inclusion of conclusive warranty disclaimers. Usually, such disclaimers can be very broad, since technology is typically made available for no cost or for a very low cost. The nonprofit organization X/Open, for instance, has included unusually broad disclaimers (compared to normal commercial agreements) in its agreements of all types.

Unfortunately, implementing some elements of such a program can be tedious. Individual member companies and their counsel must be educated as to the importance of ownership issues in the consortium context to understand why members should be expected to vary from standard, and otherwise prudent, internal practices.

A consortium also must decide whether or not to throw its standards into the public domain, or copyright them and make them freely available. Usually, the latter is the better course, since it enables the consortium to maintain better control of the standard.

Trademark protection. To truly control its standards, a consortium should undertake a rigorous program of selecting, registering, and maintaining trademarks. If certification is the group's goal, trademark registration is even more important. Trademark practice, as it applies to certification, is an unusually arcane and complex area of intellectual property law, partly because the number of consortia and other entities that have conducted activities in this area is relatively small. The range of registration choices includes the classic product trademark (for example, for a specification), service marks (for advisory services), collective membership marks (to indicate membership in the consortium), and certification marks (for certified products). There are alternative methods of registration within this spectrum, with differing advantages and strategic justifications. Whatever type of trademark the consortium selects, its use of the mark in its literature, contracts, and certification program must adhere strictly to regulations. Failure to comply can result in loss of rights in the mark.

Consortia must also consider whether or not to seek registrations in other countries—which can become a highly expensive process. For this reason, the practices of actual consortia range from merely relying on common-law first-use rights (the practice of many purely standard-setting bodies) to worldwide registration programs (the choice primarily of strategic consortia).

ALTHOUGH THEIR POPULARITY AS STRATEGIC VEHICLES may ebb and flow, the effectiveness of at least the less strategic consortia such as OMG and the X Consortium for establishment of useful standards can scarcely be seriously questioned. Otherwise, why would so many of their standards and specifications be recognized by the de jure standard-setting organizations? Nevertheless, their effectiveness is sometimes disputed in the trade press, usually because some strategic consortia have a penchant for hyping standards that are neither complete nor as effective as claimed. Likewise, some standards supported by nonstrategic consortia have lagged in both development and implementation. Finally, the much abused term *open* has far too often been applied to standards that do not facilitate any level of interoperability significant to the market.

On the other hand, many consortia can claim real work and valuable implementations. To improve upon their predecessors' records, some consortia, such as OMG, require that an implementation be commercially available at the time that a respondent to an RFT tenders a specification for adoption. Others, such as the X Consortium, develop and offer an implementation themselves. These consortia will quite certainly soldier on—to the benefit of the industry—while the flashier strategic groups rise and fall. 𝖀

Further readings

The consortia discussed in this article are of so recent vintage that little academic analysis of them has been published. The trade press and literature disseminated by the consortia themselves offer more information. The following are useful sources:

Clapes, A.L., *Softwars: The Legal Battles for Control of the Global Software Industry*, Quorum Books, Westport, Conn., 1993. This

interesting and opinionated book has an excellent chapter (pp. 261-274) on the open systems debate and the way it has played out in the consortia trenches.

The Evolution of Open Systems, 88open Consortium Ltd., San Jose, Calif., 1992, 16 pp. Good, concise review of the history and evolution of open systems, with emphasis on the current role of binary compatibility standards.

Gomez-Casseres, B., "Computers: Alliances and Industry Evolution," in *Beyond Free Trade: Firms, Governments, and Global Competition*, D.B. Yoffie, ed., Harvard Business School Press, Boston, 1993, pp. 79-128. Analyzes computer industry alliances with specific reference to RISC consortia.

The World of Standards, 88open Consortium Ltd., 2nd ed., San Jose, Calif., 1991, 243 pp. plus introduction. Excellent source book, containing detailed descriptions of over 100 important recognized standards and the organizations that sponsor them, as well as descriptions of significant products such as Windows, Motif, PostScript, SVR4, and others.

Andrew Updegrove is a partner in the Boston law firm of Lucash, Gesmer & Updegrove, which concentrates its practice in the representation of high-technology companies and consortia. He has represented or consulted with many of the largest consortia, including Object Management Group, X Consortium, 88open, Interactive Multimedia Association, and CAD Framework Initiative, to name but a few.

Address questions and comments about this article to Andrew Updegrove, Lucash, Gesmer & Updegrove, Attorneys at Law, One McKinley Square, Boston, MA 02109; 73314,3215@compuserve.com.

Reader Interest Survey

Indicate your interest in this article by circling the appropriate number on the Reader Service Card.

Low 165   Medium 166   High 167

The Development of ATM Standards and Technology: A Retrospective

The telecommunications and information technology industries currently see asynchronous transfer mode as the next major infrastructure technology. With roots in experimental switching technologies, ATM evolved along lines suggested by standards organizations. Because its services and technology coalesced in such organizations, it has achieved global acceptance. Though now finding use in private and public networks, ATM must still overcome outstanding technical, economic, and regulatory issues before it becomes a major commercial, and therefore standards, success.

Richard Vickers

Northern Telecom

The telecommunications and information technology industries currently see asynchronous transfer mode for broadband telecommunications support of multimedia services as the next major infrastructure technology. Although it had its roots in both technology and standards development, public standards bodies provided the focus for technological development of ATM. Though claims for ATM's commercial success are premature, the current industry momentum behind the technology is immense.

It is interesting to recount how ATM evolved, what challenges developers of its technology and standards faced, and how they addressed these challenges. A number of hurdles remain, including regulatory, tariff, economic, and technical challenges. Paradoxically, ATM might fail in the short term because of its current success; expectations may be too high, and the short-term demands may be greater than the technology can deliver. However, no other technologies appear to match the potential of ATM for supporting the same variety of applications in local-area, wide-area, private, and public networks.

ATM development has passed through three main stages. The first stage started with a service vision, a technology vision, and a standards base. This first stage culminated in the issue of the first framework CCITT (Comité Consultatif Interna-

tional de Téléphonique et Télégraphique) Recommendations in 1988. The core ATM development occurred in industrial and university research laboratories, and in two or three public network standards groups that provided the catalyst for the convergence of protocols. This second stage involved development of a set of implementable base standards. Since completion of the base standards, the technology has moved into the implementation, or third, stage. A much wider community of interest is now addressing both standards and technology: customer-premises equipment (CPE) and terminal equipment manufacturers have joined the original public network carriers, equipment manufacturers, and researchers.

As public network standards bodies carried out the majority of front-end ATM development, our discussion here should start with a survey of the key organizations and their operating methods. Then we can move onto the development of ATM itself. We will look at early activities in terms of standards and technology developments. The coalescence of these activities into a single stream occurred exclusively in the formal standards development process. For the first time, a single global standard for a transport technology arose. Subsequent to the formal standards-setting process, the technology moved into different working environments, and into differ-

ent applications. End users are just now becoming involved, but here we must carefully divorce the technology from its application. Technology is of little interest to end users; applications butter their bread.

The standards-setting process

The first stage of the standards cycle is the development of the standard. A standard is generally not sufficient to define equipment or an interface. Specifications use standards as key components, and will define equipment and interfaces to a level that product developers can implement them. The specification phase translates as the application of the standard. The final responsibility of the standard is to ensure interoperability between different implementations of the same standard.

Standards development. In the United States, Committee T1, an American National Standards Institute-accredited body, generates public network telecommunications standards. In Europe, the European Telecommunications Standards Institute develops related standards. Other countries also have national standards bodies. Internationally, the standards take shape as Recommendations by the Telecommunications Standardization Sector of the International Telecommunications Union (ITU-T—formerly CCITT), the ITU being a treaty organization between the national governments of the member countries.

Each of these standards bodies works on the development of standards based on written contributions against Project Proposals (T1) or Study Questions (ITU-T), reaching agreement on a consensus basis. A vote of the members tests the formal consensus. The consensus process does not necessarily mean unanimity in the agreement of the standard; it does ensure that any dissenters have their objections heard and discussed in open forum. In practice, most organizations work diligently to clear comments and objections before issuing the standard. In Committee T1, for example, external ANSI review follows two stages of balloting by members and comment resolution.

Although in the past each of the national or regional organizations has tended to produce independent national standards, a trend to place the standards development effort into the international forum has now developed. National standards refer directly to the international recommendations together with any national enhancements and option selections. Using the ITU-T Recommendations ensures that a designer does not face ambiguity because two different texts describe the same requirement, and also avoids unnecessary duplication of editorial effort in the development of the text.

The national or regional standards bodies perform most of the base technical work before passing it to the ITU by means of written contributions. The transfer is not generally direct, because the initiating standards body is usually a private-sector organization, but delegations to the ITU are

---

The final responsibility of the standard is to ensure interoperability between different implementations of the same standard.

---

responsible to their national governments or telecommunications administrations. Before they can be considered as representing the national position, generated contributions to the ITU must be endorsed by the appropriate committee, with the endorsement process generally performed in national ITU bodies.

For example, in the United States, contributions generated by Committee T1 go to the appropriate US Study Group. The US Study Groups have the delegated authority of the Department of State to establish US positions. Individual organizations can have membership in the ITU either in the Recognized Operating Administration category if they are a carrier, or Scientific and Industrial Organization if the organization is research oriented or a manufacturer. Although the process seems unwieldly, it is mitigated by the fact that most delegates attending the ITU also work in national standards development.

At the beginning of the standardization activity on ATM, single working groups in both Committee T1, ETSI, and CCITT undertook most of the work. With diversification of the technical work, the formal standards development process has expanded from the original CCITT Task Group and Committee T1 Subworking Group until it now spans a number of study groups and T1 working groups. This diversification, however, introduced a communications problem between the various activities. Both the ITU and Committee T1 have a formal process of liaisons between the various groups. The formal communication proceeds with Liaison Letters between the separate groups. A Liaison Rapporteur, who generally attends the meetings of the two or more groups involved in the liaison, usually supplements these letters.

Standards application. The individual standards developed by the standards bodies are insufficient in themselves for a designer to build equipment. The published standards provide the essential components for use in developing formal specifications. To build equipment, the designer must select an appropriate suite of standards and tailor it to the application. The application of the standard to an equipment specification may require some optimization, such as elimi-

The core standards activities are diversifying from the original development of the ATM protocol, to cover items such as signaling, network management, and equipment specification.

nation of options, and perhaps some customization to provide features not envisaged when the standard was originally developed. The formal standards development bodies do perform some applications work. For example, the protocols are generally developed as a coherent stack. Much of the work of specification, though, takes place outside the formal standards development process.

To move the standards to the implementation and specification stage, a new body entered the arena of ATM—the ATM Forum. Founded in October 1991 by Northern Telecom, Sprint, Cisco, and Adaptive, it is developing implementation agreements, which provide the selection of standards and options within standards to form a particular interface. The ATM Forum also has a stronger customer-premises equipment vendor community than the formal standards bodies, and is extending the formal standards to encompass the requirements of that community.

The ATM Forum reaches decisions by a majority vote, rather than consensus. This procedure allows a more rapid closure of issues, but it does risk alienating parts of its membership. Other than the ATM Forum, specifications tend to be customer-specific. For Regional Bell Operating Companies, Bellcore provides the development of specifications in the form of Technical Requirements and Technical Advisories. Although they have been developed for the RBOCs, these documents will probably form the basis of requirements for most US carriers.

Conformance and interoperability. We achieve interoperability when different implementations of the same specification correctly provide the intended service or application. This requires that individual protocols will work with the protocols above and below in the protocol stack. In reality, interoperability comes in two phases. First, conformance testing measures an implemented protocol against the standard. Second, interoperability testing checks that when two imple-

mentations of the protocol interwork, the system as a whole meets its functional requirements. Currently, the interoperability aspects of standardization comprise a relatively minor portion of the total standards effort. However, conformance and interoperability will become increasingly important issues as more vendors enter the arenas of ATM and B-ISDN (Broadband Integrated Services Digital Network).

Many standards under development now contain a protocol implementation conformance statement (PICS) pro forma. The PICS consists of a checklist of requirements (both mandatory and optional) contained within the standard. A manufacturer can indicate the level of conformance of an implementation by using the PICS pro forma and checking off requirements on an item-by-item basis. The result is the PICS for a particular implementation.

Complementing the PICS pro forma is an abstract test suite, consisting of a set of test cases that rigorously check the behavior of a protocol implementation through all possible state transitions of the protocol. The abstract test suite has traditionally been generated outside the formal standards development bodies, but recently the International Standards Organization/International Electrotechnical Committee (ISO/IEC) and the ITU-T have been developing abstract test suites for some protocols. Interoperability testing tends to be less rigorous that conformance testing. Typically, it involves observing the behavior of two or more interconnected implementations. Generally in interoperability testing, we test the normal operating conditions of the protocol, but not all of the error legs. Conformance testing and interoperability testing are complementary steps in ensuring interoperability. The ATM Forum has recently established a Testing Subworking Group to address interoperability as part of its mandate, and is currently defining its work plan.

Other ATM activities. There are other bodies with an interest in ATM. For example, the Internet Engineering Task Force is working on the use of transmission control protocol/Internet protocol over ATM. There is also work in the International Standards Organization on ATM rings. ATM has now moved beyond the formal standards-setting process to the applications over ATM. The core standards activities are diversifying from the original development of the ATM protocol, to cover items such as signaling, network management, and equipment specification.

"Fast" switching—the technology roots

In the early 1980s, two basic digital switching technologies existed: 64-Kbps circuit switching and X.25 packet switching. The circuit switches were actually voiceband switches that used digital crosspoints; the idea of end-to-end digital circuit switched service was under development, but not yet a reality. Only a very limited number of such switches were in service in the 1980s. X.25 was a packet switching technology for low-speed data applications. It used heavy-

weight, link-by-link flow control and retransmission protocols to avoid loss from congestion and to recover from errors induced by the underlying transmission systems of the time, primarily copper-based with some digital radio systems. The flow control procedures allowed operation at high network efficiencies on the basis that the cost of transmission was high. X.25-based networks provided very high quality and reliability of data transfer. While they still fill a vital need in the business of reliable data transfer, their application is limited to low-speed data.

The telecommunications industry had traditionally introduced new services to operate over a voiceband channel, and had been quite successful in doing so. X.25 required an overlay network approach; dedicated transmission, switching, and operations support had to be provided. The network components were expensive, and the traffic volumes small, with the result that X.25 networks struggled for a long time to become profitable.

The poor economics associated with providing overlay networks for each new service offering clearly pointed to the requirement for a different approach. The approach now taken is to provide a single switching technique and network capable of supporting a wide variety of services. The commercial success of the network would not then hinge on the commercial success of a single service. The research of the early 1980s therefore aimed at developing much more versatile switching techniques than either the conventional circuit switch, or the X.25 packet switch, since its ultimate objective was a single, or integrated, network for all services.

Three different techniques were under development at the time: fast circuit switching, fast packet switching, and asynchronous TDM. These techniques generally relied on some form of preprocessing for routing and in-band labels associated with the data for switching. The three systems differed in their method of multiplexing. Fast circuit switching used conventional TDM position multiplexing. Fast packet switching used variable length frames carried on virtual channels. The header of each frame carried virtual channel identifiers. Asynchronous TDM was a hybrid of the two previous schemes. It used fixed-length frames on virtual channels, and carried labels in the header of each frame. Figure 1 illustrates the different multiplexing methods.

All three of these techniques minimized the processing required for switching. Fast circuit switching retained the characteristics of conventional synchronous TDM switching
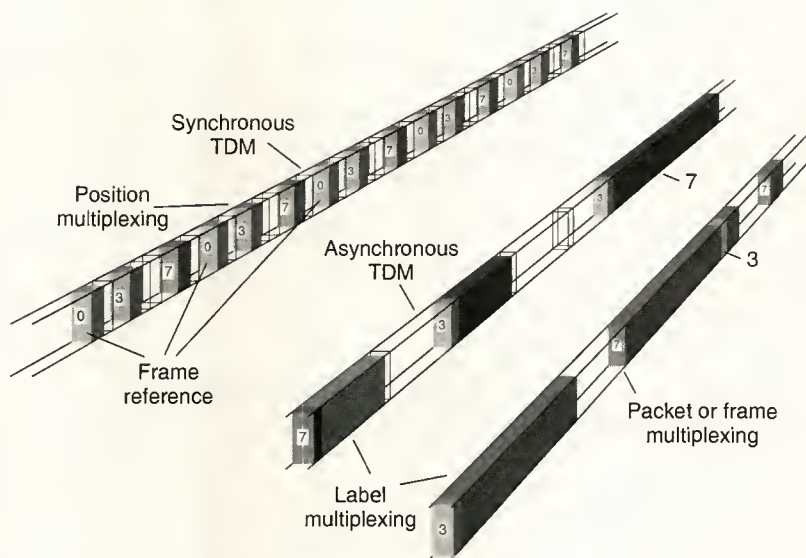
Figure 1. Multiplexing methods. Numbers indicate channel or virtual channel.

systems in that the bandwidth of a channel was fixed, resulting in a lack of flexibility to accommodate different source bit rates. For this reason the technology was not pursued, and especially because most sources (including voice and video) do not intrinsically generate constant bit rates.

Researchers developed the fast packet technology, but for the transfer of medium-speed data only, rather than for the mix of voice and data. This development produced frame relay, now deployed by a number of carriers. Frame relay will likely serve as an intermediate technology to support applications such as LAN-to-LAN traffic before full ATM service becomes available. Somewhat ironically, current plans include using ATM to support frame relay.

As asynchronous TDM development emerged, it became known as ATM. The early versions were somewhat primitive, and required considerable development, but all the basic principles were in place. The label multiplexing and the fixed-length packets, now called cells, were all components of the original asynchronous TDM technology.

These experiments set the technology base for what was to be developed in the standards. They established the viability of using hardware and in-band headers for controlling switching. The technologies themselves were experimental, either in the form of laboratory prototypes, or limited field trials, but they did provide one of the cornerstones for what was to follow in the development of ATM.

Early standardization efforts

The major standardization effort of the early part of the
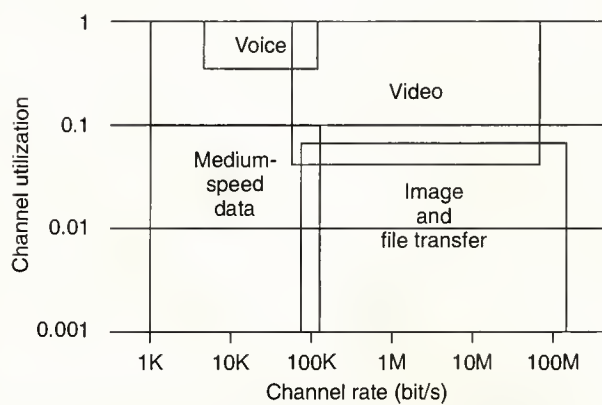
Figure 2. Bandwidth/channel utilization characteristics.

1980s was on the ISDN. Intended to provide an end-to-end digital service, both circuit switched (64 Kbps) and X.25 packet switched, ISDN concentrated on integrated access from the transport perspective, but based on existing digital switching technologies. It made no attempt to coalesce the underlying switching structures.

In 1985, industry recognized that new services would emerge that would require transport capabilities beyond those supported by the ISDN then under development. The term B-ISDN then came into existence, meaning broadband aspects of the ISDN. Note that B-ISDN was considered a part of the ISDN.

B-ISDN split from the mainstream of ISDN development in 1986; both CCITT and ANSI created separate subgroups. In the longer term, this separation of the work allowed for a less constrained development of the standards.

During the early phases of the standards development, most of the work concentrated on the services that would use B-ISDN, but with some discussion of the supporting transport technologies. The issue of the first CCITT B-ISDN recommendation in 1988[1] marked the completion of the early phase of development. This recommendation set the framework for subsequent developments.

B-ISDN is an entire network concept. It encompasses transport, multiplexing, and signaling, as well as operations, administration, and maintenance.

Transport. By the early 1980s, we had optimized the transmission hierarchy around the capabilities of copper and radio-based systems with limited bandwidth capabilities. We generally timed transmission systems from local free-running clocks. To accommodate timing differences, we performed multiplexing above the primary rate (DS1—1.544 Mbps) asynchronously by bit stuffing techniques. At the same time, optical fiber systems were starting their deployment in volume. A single optical fiber has enormous intrinsic bandwidth,

but the transmission hierarchy of the time was designed to handle relatively low bandwidth systems, and was really not suited to efficiently manage the large bandwidths available with optical fiber systems.

To extract the most from fiber transmission, researchers developed a new synchronous transmission format. In North America it is known as Sonet (synchronous optical network) and in Europe as SDH (synchronous digital hierarchy). A detailed technical discussion of Sonet/SDH is beyond our scope here, but it possesses two extremely important properties for B-ISDN. First, bandwidth can be easily concatenated to form very large channels. Second, it provides a transport capability and common management platform that is independent of the format of the payload. Both these properties are important for B-ISDN and provide a strong transport base.

Transfer modes and multiplexing. Sonet provided the transport platform suitable for carrying broadband services. Work in standards now concentrated on the development of suitable multiplexing methods for carrying a number of services with different bandwidth and other characteristics (see Figure 2). Three modes of transport were under discussion: synchronous transfer mode, ATM, and packet transfer mode. STM is the traditional digital voice transfer mode, where the transfer takes place with fixed blocks of information at fixed intervals. Packet transfer mode proceeds using variable-size blocks of information at varying intervals. ATM uses fixed blocks of information at variable intervals.

Complementary to the discussions on transfer mode were the talks on multiplexing. Because of the difficulties of carrying constant bit rate information and the asynchronous nature of label processing in such systems, pure packet multiplexing eliminated itself fairly quickly. Hybrid circuit/packet structures and asynchronous TDM remained for considerations. The proponents of the hybrid system based their arguments on compatibility and interworking with existing systems, whereas the proponents of asynchronous TDM cited the flexibility of the system and its ability to automatically adapt to variations in traffic mix.

After much debate during 1986, the standards bodies agreed to focus on studies and standardization of ATM, and the standardization of asynchronous TDM as the multiplexing technique followed almost automatically.[1] The reasons for the decision were complex, but one major advantage was that we could now employ a single multiplexing layer for all services on top of the common transport layer of Sonet/SDH, thus simplifying layer management functions. Note also that the overall ATM system efficiency will generally be greater for a mix of services than a circuit-switched system despite the added overhead because of ATM's ability to rate adapt the transport to the requirements of the source. The ability to statistically multiplex the resultant variable bit rate sources may further enhance ATM's efficiency.

Other standardization activities. Two other major stan-

dardization activities occurred through the 1980s: fiber distributed data interface LAN activities (in ANSI X3T9) and the metropolitan area network activities (in IEEE P802.6).

The Fiber Digital Data Interface project was initiated in the early 1980s to provide an optical fiber-based LAN at 100 Mbps. Envisioned at its inception as the interconnection of mainframe and peripherals, the standard was fully developed and products are now available. FDDI now tends to serve as a backbone LAN rather than the original, envisaged application. Developed purely for data applications, a subgroup of FDDI spun off early in the development cycle to look at the carriage of isochronous services. This subgroup defined a new signal structure (FDDI-II), where we could divide the 100-Mbps bandwidth between data and isochronous applications. This hybrid packet/circuit structure had obvious application in multimedia LANs, but in retrospect the timing was wrong, predating the market requirements for multimedia LANs by some years.

FDDI is deployed in LANs today. In addition, some of the technology developed for FDDI is finding application in other areas. Of interest to ATM, the ATM Forum has defined an ATM user-to-network interface for customer premises applications using the physical layer of FDDI.

Activity started for the MAN standard in the early 1980s, but languished for some time, until in 1987 a protocol (distributed queue dual bus—DQDB) was proposed that had adequate bandwidth/distance characteristics for MAN applications. Also, its format closely resembled the format under development for B-ISDN, in that it was based on fixed-length packets, or cells. Because the format was so similar to the ATM cell format, investigators expended considerable effort to maintain commonality between the cell sizes in ATM and IEEE P802.6. Work on the basic IEEE P802.6 standard concluded in 1990.[2]

IEEE P802.6, like FDDI-II, has an isochronous capability. In IEEE P802.6, cells placed at fixed intervals provide the isochronous capability, giving it built-in multimedia capabilities. IEEE P802.6 has major importance in the transition of B-ISDN services. We use it as the access technology to support the switched multimegabit data service defined by Bellcore. SMDS is a broadband connectionless data service with access rates initially at the DS1 (1.544 Mbps) and DS3 (44.736 Mbps) rates. Initially, we thought that the service would be supported on an overlay IEEE P802.6 network; more recently the network support for the service has moved to an ATM base.

Development of the ATM standard

Once the decision was made to use ATM during 1986, the parameters of the ATM cell had to be standardized, which led to protracted and heated discussions in the standards forums. The North American position centered on a 64-octet cell payload; the European position centered on a 32-octet payload.

> *Once the decision was made to use ATM during 1986, the parameters of the ATM cell had to be standardized, which led to some protracted and heated discussions in the standards forums.*

In North America, the IEEE P802.6 committee had also adopted a cell-based format for MANs, as well as a particular format of a 64-octet payload and a 5-octet header. There was also SMDS to consider. SMDS is a datagram service that routes on an E.164 address[3] embedded in the header. The header itself could fit into one cell (32 or 64 octets), but the routing decision had to be made very quickly. Processing the header in one cell time would eliminate queuing of headers. With a 64-octet payload, observers felt that there was sufficient time to perform the processing, but 32 octets left insufficient time. In addition, a 32-octet payload with a 5-octet header would result in a maximum efficiency of 86 percent, considered too low, especially when combined with the efficiency of Sonet/SDH, and detraction of the 32-octet payload by adaptation layer functions.

Europeans made a strong push for a total integration of services, including voice. For performance reasons the delay requirements on voice are quite stringent. At 64 Kbps it would take 8 ms to fill a 64 octet cell, and 8 ms from the European perspective was unacceptable, as it would mean major investments in the deployment of echo control devices to counter the additional delay. Europe was therefore proposing a cell size of 32 octets.

After much debate, the various parties agreed upon a cell consisting of a 48-octet payload and a 5-byte header in June 1989. The selection of a common cell was vital to the development of ATM. The adoption of different cell sizes in different regions of the world would have prevented interworking at the ATM layer. Instead, the signals would need reconstruction in their original form (64 Kbps, data packets) and redivision into the different cell structure of the receiving country. This form of interworking would have to be performed on a service-by-service basis, and in the long term would have severely constrained the development of international ATM networks. The compromise cell size is

CLP	Cell loss priority
GFC	Generic flow control
HEC	Header error check
PTI	Payload type indicator
VPI	Virtual path indentifier
VCI	Virtual channel identifier
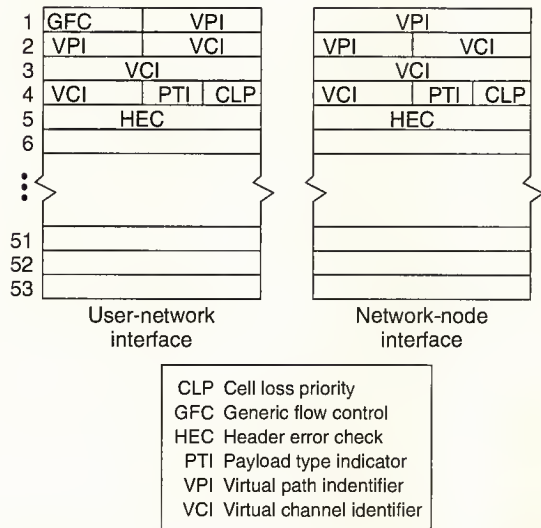
Figure 3. ATM cell formats.

probably not ideal, but the need for international compatibility outweighed the technical arguments put forward for different cell sizes. The IEEE P802.6 Committee subsequently modified its cell size to align with the 48-octet payload.

The parties also agreed on mapping of ATM into the 155-Mbps Sonet/SDH payloads. The combination of these mappings, and a common cell format meant that for the first time we had a common transport format that we could apply in all regions of the globe.

Conferees in subsequent meetings also accepted and developed other functions and fields in the header, until the final header and cell formats (Figure 3) were incorporated into the 1992 versions of the CCITT ATM layer recommendations.[4,5] The formats for use at the user-to-network interface and for use internal to the network (network node interface) differ in one significant detail. At the UNI the header dedicates four of its bits to a function called generic flow control. At the NNI it allocates these bits to the virtual path identifier.

It is interesting in looking at these header structures to examine the developments from the primitive two-octet header implementations of some early ATM applications. These developments have all arisen because of the public technical debates that occur in the standards forums, and in every sense make ATM a more viable and flexible technology.

The essence of the header is in the routing field. The routing field is the label, and provides the association of the cell with a particular channel or circuit. It is simply a number that has significance only on the local link. As the cell traverses a switching point the number will generally change, just as in a synchronous TDM system the timeslot number will generally change as the channel traverses a switch.

The routing field is divided into two portions: the VPI and the virtual channel identifier. The VPI allows us to group VCIs and switch the group as an entity. Grouped switching, which operates as the ATM equivalent of trunk grouping, was initally proposed by Australia and Japan. It also gives users access to blocks of channels, for example, to interconnect ATM PBXs. The VCI is the single-channel identifier, and thus identifies the basic switched channel.

Agreement on the size of the routing field proved elusive. The size of the routing field dictates the maximum number of virtual channels that the same multiplexed interface can support simultaneously. Again, the sizes finally selected represented a compromise. On one hand, some emerging applications of virtual path would require large numbers of virtual paths both within the network and at the UNI. On the other hand, the routing field would need to support a large VCI field for the general user of B-ISDN. The compromise probably fully meets the needs of both groups in practice. The maximum number of virtual channels that the interface can support is 16 million at the UNI and 256 million at the NNI.

The second field is the payload type indicator. Early developmental stages showed the ATM protocol would need additional capabilities to provide for operational functions not related to the user information. Unlike statistical time-division multiplexing, a virtual channel may have a very intermittent flow of user information. To check whether the channel is still viable during a period of inactivity, we could insert a dummy cell to check that it is transported over the entire virtual connection. Although it would carry the same VPI/VCI as a cell carrying user data, it is not actually carrying user data; the payload consists of information used by the network in checking the viability of the channel. The PTI allows that cell to be extracted by the monitoring function from the user data stream; if it were interpreted as user data it would cause an error.

We have found other uses for the PTI, which has been extended from the two bits originally allocated to three bits. The functions now include an in-channel multiplexing identifier and a congestion indication (both for user data), two types of operations administration and maintainence cell, and a resource management cell (for network use).

The last functional bit in the header is the cell loss priority bit. This bit provides the network with a selective discard capability. The user may mark cells high or low priority. In some cases, however, the network may change the value of the CLP bit.

The final field in the header is the header error check field. Added to prevent errors in the header that cause a misrouting of the cell into another user's data stream, this field contains the result of the CRC-8 calculation across the previous four octets. When a switch or terminal terminates the header, errors will be detected with a high probability. If the error

is a single-bit one, it may be corrected. If the error is not correctable, the whole cell must be discarded. Considerable work on the theoretical characteristics of the error check algorithms led to agreement on the detection and correction capabilities, with the results being mapped onto the measured performance of optical fiber transmission systems in the field. The ability to correct single bit errors improves system performance, but with a slight degradation of the error detection capability. Note, though, that certain types of transmission systems use block coding schemes. Such schemes introduce error multiplication. If we use ATM over such transmission systems, we should use the header error check only in detection mode.

The GFC field at the UNI is the only field where we have not fully defined the functions and protocol. The prime purpose is to provide arbitration between contending terminals on the same interface. Standardization of GFC should be completed sometime in 1994. Note that for a simple uncontrolled interface, the GFC field is not used. Therefore, absence of complete encoding does not impede the deployment of ATM.

That completes a brief description of the ATM layer fields and functions. With the exception of the GFC, the ATM layer standards were completed in 1992. ATM is a compromise technology, and as such it retains some of the features of the parent technologies and loses others. As a result, to carry specific services over ATM, we must restore certain functions to the overall protocol stack before the service can be carried. The adaptation layers carry these additional functions.

Adaptation layers

By design, the transmission and multiplexing functions are largely service generic; that is, these layers are not customized for service-specific functions. The adaptation layers provide the functions that enable specific services to be carried. While a full description on the adaptation layers is beyond our scope here, a basic understanding of the type of functions they provide will be helpful. For a full description, see the CCITT recommendations I.362[6] and the protocols described in I.363.[7]

The early work on services established classifications based on various service characteristics (Figure 4). These service characteristics identified requirements that the ATM layer did not necessarily provide. To provide these characteristics to the services, we identified four adaptation layer types, labeled Types 1 to 4:

- Type 1 adaptation for constant bit rate services (Class A). requires provision for reconstructing the service clock at the destination. We have developed a protocol for Type 1 adaptation. This protocol will serve initially in the carriage of services that emulate the performance of an existing constant bit rate service, such as DS1. Providing

	Class A	Class B	Class C	Class D
Timing between source and destination	Required		Not required	
Bit rate	Constant	Variable		
Connection mode	Connection-oriented			Connection-less

Figure 4. Classification of services.

the adaptation layer functions requires one of the 48 payload octets in each cell.
- Type 2 adaptation is for variable bit rate services where the timing relationship between the cells is significant to the service (Class B). One service that will require this type of adaptation is variable bit rate video. However, we have not yet defined VBR video services, and the adaptation layer protocol is not yet required.
- Type 3 adaptation is for VBR connection-oriented data, where the timing relationship between cells is not important (Class C). It provides for the basic segmentation of variable length frames into cells at the sending end, and the reassembly of the frame at the receiving end. It also has error checking and a multiplexing capability built into each cell. The adaptation layer functions consume four of the 48 payload octets.
- Type 4 adaptation is for connectionless data (Class D), and is almost identical to Type 3. The only difference is in the use of the multiplexing identifier. Type 3 assigns a value for the duration of the connection whereas Type 4 assigns the value only for the duration of the frame.

Subsequent to the identification of these four adaptation layer types, a new simpler adaptation layer emerged for connection-oriented VBR data service. The Type 3/4 adaptation layer requires that four octets of information be detracted from the cell payload to carry necessary information from transmitter to receiver. The Type 5 adaptation layer does not use any of the 48-octet payload for adaptation, but instead requires an identifier in the ATM header, and uses the in-channel multiplexing identifier function of the payload type field. Type 5 provides only segmentation and reassembly at the cell level. It provides all other functions, such as length checking and error detection, on a frame basis.

The introduction of Type 5 has virtually made Type 3 obsolete. Services such as signaling and frame relay have moved from Type 3 support to Type 5. Type 5 represented a challenge to the standards community in two ways. Type 5 originated in the customer-premises equipment industry, a group

> **The thrust of standards development has moved away from ATM itself to other standards required for full B–ISDN capabilities.**

that works on much shorter time scales than the public network industry. The challenge was to accept a proposal from a different segment of the industry and complete the standardization much more quickly than is normal for such standards. The standards community met most of these challenges, and Type 5 AAL achieved international agreement within eighteen months of its first appearing in Committee T1.

The work on adaptation layers will continue for some time. As new applications emerge, we will likely define new adaptation layers. However, the development of new adaptation layers will have minimum impact on public network facilities. Since adaptation does not form part of the base transport, we implement it in the CPE or in network servers. The core ATM transport will not be affected.

Developments since 1992

With completion of the base standards in 1992, ATM entered the implementation phase. Until then public network operators and equipment had almost exclusively developed the standards. They had also been developed to a fairly relaxed timetable, as there was no real commitment (apart from trials) to the implementation of networks based on the standard.

During 1991, some organizations in the business of producing high-end workstations and other CPE equipment realized that ATM technology could fill some of their needs. Workstation I/O capabilities were reaching 40 Mbps. At those rates the then current state-of-the-art LAN (FDDI) could handle only a total bandwidth of 100 Mbps, a capacity that would not be suitable for networking such high-speed applications. Multimedia was also emerging, with trials of multimedia editing workstations operating on voice, video, graphics, and text files. The development environment for ATM standards had very quickly become much more pragmatic.

Most immediately, the CPE industry needed Implementation Agreements for interfaces to terminals and ATM switches. Such agreements would provide a complete definition of an interface, including operations, management, and signaling. Standards provided the building blocks and some guidelines on protocol stacks to support certain applications.

Announced in October 1991 with the specific charter of producing such Implementation Agreements, the ATM Forum has issued such agreements on the UNI.[8] Besides using formal standards, in certain instances it has also filled specific needs by developing extensions to the standards (such as for multipoint connection capabilities) or new technologies where no standards exist (such as ATM over unshielded twisted-pair cable). Eventually, some of these developments should be fed back into the standards process.

Formal standards development since 1992 has also become more pragmatic. Recognizing that 150 Mbps could not be economically deployed to all customer premises in the near term, the standards community is developing interfaces at 44 Mbps and 1.5 Mbps, based on the old, but widely available asynchronous transmission formats. The thrust of standards development has moved away from ATM itself to other standards required for full B-ISDN capabilities: signaling, OAM, and interworking with other networks.

The standards-setting process

The development of public network standards has traditionally been slow. The ITU used to work on a four-year cycle for issuing new recommendations. In periods of slow technological change this four-year cycle was adequate, but the current environment, with topics such as ATM and related standards, requires a much faster development cycle. The standards community is establishing a number of processes to aid this more rapid development.

The ITU has decoupled the approval of Recommendations from the four-year study period, and now has a process to generate new Study Questions during the study period. Its formal approval process is still slow, taking about a year from the completion of the text to formal approval of the standard, while publishing the Recommendations adds more time. The ITU is now developing electronic publishing and document handling capabilities, which should speed the publishing process.

The need to publish the formal documents in three languages is another expensive and time-consuming processing issue within the ITU. Currently, people perform the translations, with technical experts reviewing the translated text for consistency with the original. Before long, computers may be able to perform the initial translation, although the technical review of accuracy will still have to be manual. Again, this will speed the standards development cycle.

In the early days of ATM standards, progress was slow, due largely to misunderstandings between the technical experts of different countries that resulted in entrenched positions and dogma. The debating environment was formal and was not conducive to informal technical discussions between participants. To provide a more conducive environment, representatives of Committee T1, ETSI, and TTC (Telecommunications Technology Committee) of Japan

agreed that an annual B-ISDN Technical Workshop should be established. Held in 1991, the first workshop discussed specific issues in an informal atmosphere to help eliminate conflict resulting from technical misunderstanding. In this objective, the workshop has been extremely successful.

Standards, implementation, and interoperability. The underlying purpose of standards is to ensure that when various standardized components are brought together that they interoperate. Standards also should define boundary conditions, but are not intended to constrain implementation. However, as protocols become more complex, interoperability becomes more difficult to write into the standard. There are a number of reasons why this is occurring.

1. *The use of prose text in standards.* Using prose text makes it very difficult to define a function precisely in such a way that it can be interpreted unambiguously, or such that two different implementations of the same function will interwork.
2. *The definition of options within standards.* Options may be included in standards for a number of reasons, some technical, some political. For whatever reasons they exist, options can lead to confusion and incorrect implementations.
3. *Errors of omission.* Prior to implementation, verifying the completeness of a standard is difficult. When the standard is implemented, different designers may choose to implement the missing parts in different ways, leading to interoperability problems.

The use of prose text in standards definition is traditional. However, the inexact meanings of prose text cause problems, even before standards are issued. Standards developers spend much time debating whether the wording reflects the intent of the standard, and whether it is unambiguous. But even agreement may not eliminate all problems. For a direct comparison, look at the law, where even more precise meanings to language do not entirely eliminate problems of ambiguity and interpretation. Therefore, case law must supplement legislation to provide judicial interpretation of the legislation.

In the field of standards, replacing prose description with formal description techniques has helped. FDTs provide a method of specifying protocols or behavior unambiguously and precisely. Development of FDTs has proceeded for a number of years, and in the case of SDL (functional specification and description language) the technique has been used in standards for some time. However, the SDL portion of the standards is only informative, one reason being that SDL is not sufficiently abstract to be independent of implementation. Also, from an accuracy viewpoint, we have not had the tools to debug the SDL as written.

SDL is just like a computer language; specifications written in SDL must be verified for semantics and syntax, then run to check that the protocol as defined meets its original requirements. Only recently have tools to perform these tasks become available on the open market. SDL is only one of several FDTs under development, but it is certainly the most mature and widely used. Other FDTs may be more independent of implementation, but are not as easy to use, nor do the tool sets exist to assist the designer.

If FDTs are to form the normative part of a standard, the standards community will have to reconsider the sanctity of the principle that standards should not constrain implementation. One possibility is the use of the tool in reference implementations. Even though a designer would have freedom of implementation, the validity of the design would always be judged against the reference implementation.

The standards development bodies are just starting to work on conformance and interoperability. As systems become more complex, conformance and interoperability will have to form a much more important role in the standards development process. The standards bodies cannot claim to be doing a complete job until they recognize the importance of these topics.

The future of ATM

All segments of the telecommunications industry now believe that ATM will be the transport technology for cost-effectively supporting multimedia services. However, before it fulfills its promise as the ubiquitous transport for broadband services, it must still face a number of challenges, some technical, but some legal, regulatory, and business related. These factors will determine where the technology has its first applications, and how it will develop.

Technical issues. The standards completed to date are sufficient for the implementation of basic services on ATM. For certain types of service, though, the network efficiency could be quite low, and these are the highly bursty data services. Within the network, using large trunk cross sections (600-2,400 Mbps) will allow the law of large numbers to come to bear for statistical multiplexing of services with peak bit rates up to a few tens of Mbps. The local-area and private networks, however, generally have an insufficient volume of users or traffic to use the law of large numbers; such networks are frequently, indeed routinely, congested. In local networks, therefore, media access control protocols form the basis of the LANs such as Ethernet (CSMA/CD), Token Bus, and Token Ring. Such protocols work in LANs because the geographical distance is limited. More important, the LAN effectively forms a single-stage switch, making contention an easier problem to solve than in a large, wide-area network. Such protocols effectively share the available bandwidth among the users during times of congestion, rather than discarding excess traffic at the expense of degraded throughput performance. Currently defined

ATM protocols would be forced to discard should congestion occur.

The challenge is to extend the same service provided by a LAN in the local area across the wide area. Today, we interconnect LANs across a wide area by using routers attached to the LANs and interconnected by fixed low-bandwidth facilities. The same method applies to an ATM network, although the ATM network can provide higher bandwidth pipes. The evolution of the network will see provision of LAN-like capabilities provided on a wide area, ubiquitously available basis. Standards bodies are currently discussing a number of mechanisms, including Fast Reservation, link-by-link flow control, and backward congestion notification.

Other major technical areas under debate include signaling and network management. ITU-T Study Group 11 is developing signaling protocol. Signaling work focuses on the capabilities that B-ISDN will provide to meet marketplace requirements. The group has scheduled two releases:

- Capability Set 1, scheduled for completion by the end of 1993, provides for basic call establishment for point-to-point connections; and
- Capability Set 2, scheduled for completion at the end of 1994, will provide for limited multipoint and multimedia capabilities.

The study group has identified some capabilities for Capability Set 3, but has not yet defined its completion. Capability Set 3 will provide for more sophisticated multimedia and multipoint capabilities.

Development of network management is progressing along two streams. For private networks, the customer network management in North America is evolving from the Simple Network Management Protocol defined by the Internet Engineering Task Force. In the public network, the Network Management Protocols build on OSI, the ISO/ITU-T protocols stack. In Europe, many administrations favor the OSI approach for customer network management.

The standards for deploying a basic ATM network are complete. Such a basic network would have only virtual private line capability, and would not be capable of a high degree of statistical multiplexing. The ATM Forum has developed some extensions to the formal standards to allow for switched service, and a limited amount of statistical multiplexing that will assist in getting started with some more sophisticated techniques.

Signaling developments will enrich the variety of services provided by B-ISDN, and improvements in traffic management techniques will improve the efficiency of the network. Still, network operators will need a period of adjustment as they learn how to manage ATM networks and how to tailor them to increase their efficiency and utilization. We should not overextend the ATM technology in the early days.

Economics and tariff. ATM technology operates better as the bit rate of the services gets faster, and as the number of users increases. Some issues associated with traffic management manifest themselves in start-up costs for ATM networks, particularly in the public network environment.

The cost of the equipment needed to support the service dominates the cost of ATM in customer-premises applications, and the cost and performance benefits of ATM over competing technologies will determine its success there. ATM has a start-up cost disadvantage, as even in the smaller sizes, the switches tend to have switching capacities in the gigabit-per-second region, compared with the 100 Mbps of FDDI. Even if the cost per bit per second of ATM may be a fraction of the cost of FDDI, initially only a few users networking top-end workstations could justify the ATM capacities. Consequently, ATM is likely to start as a backbone technology in applications where FDDI does not provide sufficient capacity.

For connection to the desktop, ATM has a lot of competition. Its success will depend on the development of applications that really require the tens or hundreds of megabits per second that ATM can provide. The 10 Mbps provided by a dedicated or switched Ethernet connection could provide sufficient capacity for many applications, including video, and at much lower cost than ATM. However, another model states that applications will always grow to consume the available bandwidth—that is how the personal computer industry has made new demands on processing speed and storage requirements.

In public network applications the economics of ATM are more complex. The capacity of public network ATM switches in trunking applications will start at about 10 Gbps. Compared to the total data traffic carried by the network, this capacity is enormous. The interconnecting trunks will also be very large, 150 and 600 Mbps being the internationally agreed network bit rates. As with the customer premise applications, public network applications must also address a start-up issue.

The cost of public network service to the end-user manifests itself in the tariff. A full discussion of tariff strategy is beyond the scope of this article, but certain aspects of tariffs can encourage or discourage certain behavior from the user of the network. For example, within the wide area network, it is more efficient now to shape bursty traffic to a fixed peak bandwidth than to pass it through the network unshaped. For items that do not depend on network delay, we can structure the tariff to encourage the end user to shape traffic for the network to handle it in the most efficient way.

In both public and private network applications of ATM, as with any technology, the economic factors will determine the ultimate market position. The cost of the technology must be such that sufficient volume deployment arises to justify manufacturers making major investments in cost reduction. Under these circumstances, ATM could see the levels of

growth and demand exhibited by the personal computer. With the current level of industry momentum behind it, ATM is unlikely to fail completely. However, if the cost is too high, it could remain a niche technology, much as X.25 has remained a niche network technology.

Regulation, legislation, and market trends. Historically, the local exchange carriers have operated as a monopoly under Public Utility Commission regulation and federal legislation. Under the law they cannot generally provide protocol conversion, and they are restricted in the types of service they can provide. However, their monopoly now faces challenges on a number of fronts: alternate access carriers for business services, CATV operators for residential services, and cellular radio carriers for mobile services. The regulation and legislation are now becoming major constraints to competition to those companies that operate under its umbrella. To counter the threat, the LECs have achieved some success in challenging the legislation prohibiting them from providing entertainment video services in their own operating regions. They are also obtaining more and more waivers to the protocol translation prohibitions. The whole industry is moving to a competitive basis from a regulated basis, just as the interexchange business started to transition in the early 1980s.

ATM can only benefit from such competition. As a technology suitable for supporting the richness of options that service providers will want to offer to retain their competitive edge, ATM use will certainly facilitate the generation of competition in residential applications. Its application to the public network-owned local loop and CATV delivery will provide a common platform on which to compete for services.

ATM IS IN THE EARLY IMPLEMENTATION PHASE. As we gain experience from their implementation, standards should be extended and perhaps amended. The long-term commercial success of the technology will depend not only on its continued technical development, but also on economic, regulatory, and market situations. Optimistically, ATM could trigger an appetite for bandwidth growth analogous to the appetite for computing power on the desktop. Pessimistically, ATM will remain a niche technology for high-end computing and multimedia communications. With the momentum that exists behind the development of ATM and ATM products, it is unlikely to fail completely. ▥

References

1. *CCITT Recommendation I.121, Broadband Aspects of ISDN,* Comité Consulatif International de Téléphonique et Télégraphique, Geneva, 1988.

2. *P802.6 Distributed Queue Dual Bus (DQDB) Subnetwork of a Metropolitan Area Network (MAN),* IEEE, Piscataway, N.J., 1991.

3. *CCITT Recommendation E.164, Numbering Plan for the ISDN Era,* CCITT, 1988.

4. *CCITT Recommendation I.150, B-ISDN Asynchronous Transfer Mode Functional Characteristics,* CCITT, 1992.

5. *CCITT Recommendation I.361, B-ISDN ATM Layer Specification,* CCITT, 1992.

6. *CCITT Recommendation I.362, B-ISDN ATM Adaptation Layer (AAL) Functional Characteristics,* CCITT, 1992.

7. *CCITT Recommendation I.363, B-ISDN ATM Adaptation Layer (AAL) Specification,* CCITT, 1992.

8. *ATM User to Network Interface,* V. 3.0, ATM Forum, 1993.
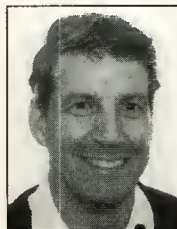
Richard Vickers has represented Northern Telecom in the development of B-ISDN standards since 1987, both in Committee T1 and CCITT (now ITU-T). Currently, he serves as chair of T1S1.5, the standards group responsible for the development for ATM and B-ISDN standards in the US, and for the development for the ITU-T. Earlier, he served as chair of the ATM Subworking Group in T1S1.5. At Marconi in Chelmsford, England, he developed digital transmission and switching equipment. For BNR in Ottawa, Canada, he developed a variety of technologies including fast circuit switching and ATM multiplexing.

Vickers graduated from the University of Wales, Cardiff, with a BSc in electrical engineering.

Direct questions concerning this article to the author at NTI, Department 6LMS242, PO Box 3511, Station C, Ottawa, Ontario K1Y 2H7, Canada; vickers@bnr.ca.

Reader Interest Survey

Indicate your interest in this article by circling the appropriate number on the Reader Service Card.

Low 168   Medium 169   High 170

A Survey of Encryption Standards

Numerous encryption standards dot the microcomputer landscape, seemingly covering every application. One nevertheless finds much common ground underlying the many standards. This survey discusses the standards and their algorithms, how they compare, how they differ, and where they're headed.

Burt Kaliski

RSA Laboratories

Cryptography is the science, or some would say the art, of secret codes. In its broadest sense cryptography addresses a number of practical problems:

- *confidentiality,* keeping messages secret;
- *origin authentication*, verifying a message's source;
- *integrity,* assuring that a message has not been modified; and
- *key management,* distributing the secret "keys" for cryptographic algorithms.

This survey focuses on encryption algorithms, the low-level, step-by-step transformations on messages that address these problems, as well as applications that involve encryption. It covers both approved standards and work in progress; the modifiers *draft* and *proposed* should help with the distinction.

Since descriptions here are at a summary level, readers seeking greater depth may refer to the standards documents or to encryption surveys such as those by Diffie,[1] Simmons,[2] which includes a reprint of Diffie's article, and Fahn,[3] which is available from RSA Laboratories or via anonymous ftp to rsa.com. Patel gives an earlier survey on security standards for the Open Systems Interconnection (OSI) reference model.[4]

Much of the encryption standards work fits into one or more security "models." The models do not specify algorithms; rather, they define services and give structures for encryption protocols. The OSI Security Architecture standard[5] is one helpful reference. Also on the road to international standardization is the Generic Upper Layers Security (GULS) standard.[6] GULS forms the basis for IEEE P802.10, a local-area network security project, and the draft ANSI X9.41,[7] a standards effort for electronic data interchange.

Many ways other than encryption exist to protect data, from access control to tamper-resistant coatings, but they are outside the scope of this article. Even in systems based on cryptography, other issues than just the codes come into play, such as random number sources and password selection guidelines. The US Department of Defense's "Orange Book" is one of many helpful references for these topics.[8]

Remember, draft standards and other works in progress are subject to change. Furthermore, with the large number of standards efforts, I may not have covered some relevant efforts. An effort's absence from this article in no way minimizes its importance.

Algorithms

An encryption algorithm is a method of transforming a message to add some cryptographic protection, such as confidentiality or integrity. Most encryption algorithms involve one or more keys, which are cryptographic variables, often

unique to one user, that control the algorithm and provide security against attackers.

Cryptographers often classify encryption algorithms according to the type of transformation and keys. Each class solves a different set of cryptographic problems. Some classes require that parties first agree on a secret key by secure means that are separate from the normal communication protocol; others do not have this limitation. I describe the algorithms standards according to one such classification: secret-key cryptosystems, public-key cryptosystems, digital signature schemes, key-agreement algorithms, cryptographic hash functions, and authentication codes. Table 1 summarizes the classes and their properties.

Secret-key cryptosystems. These algorithms encrypt and decrypt messages with a key in such a way that it is difficult to decrypt without the key. Because the encryption and decryption keys in a secret-key cryptosystem are the same, such systems are often called symmetric in the literature.

Most secret-key cryptosystems operate on messages one block at a time; a block may be 64 bits long, and the keys are usually short, say, 56 bits long. Ideally, an attacker's only approach is trial and error, which amounts, for example, to 2^{56} trials for 56-bit keys. Secret-key algorithms are generally quite fast.

Secret-key cryptosystems provide confidentiality and key management to parties who have previously agreed on a secret key. The Data Encryption Standard (DES)[9] is the primary standard. Published in 1977 and recently affirmed for a fourth five-year period, DES defines the Data Encryption Algorithm (DEA). It also specifies how to implement DEA: in hardware. Technically, software implementations of DEA, which abound, do not comply. ANSI standard X3.92[10] and Australian Standard AS2805.5[11] specify DEA.

Despite much controversy about the nature of DEA—the government never revealed its design criteria—the algorithm seems to be quite secure, as far as 56-bit algorithms go. It resists powerful attacks that have broken other systems.[12,13]

Along with DES come some standard modes of operation, including electronic codebook, cipher block chaining, cipher feedback, and output feedback.[14] These modes apply to any block cipher, not just DEA. ANSI X9.17[15] introduces the encrypt-decrypt-encrypt (EDE) mode of encryption involving two DEA keys.

Two password-based encryption algorithms defined in the intervendor public-key cryptography standard (PKCS) #5[16] are also based on DEA.

A potential new standard secret-key cryptosystem is Skip-

Table 1. Encryption algorithm classes and their properties.

Class	C	OA	I	KM	Prior
Secret-key cryptosystems	Yes	No	No	Yes	Yes
Public-key cryptosystems	Yes	No	No	Yes	No
Digital signature schemes	No	Yes	Yes	No	No
Key-agreement algorithms	Yes	Optional	No	Yes	No
Cryptographic hash functions	No	No	Yes	No	No
Authentication codes	No	Yes	Yes	No	Yes

C indicates confidentiality; OA, origin authentication; I, integrity; KM, key management.
Prior requires that parties first agree on a secret key.

jack, a classified part of the proposed escrowed encryption standard.[17] A panel of cryptography experts recently certified Skipjack, with 80-bit keys, as appearing secure,[18] but its details remain unpublished.

Secret-key cryptosystems are rarely standardized; some standards bodies explicitly omit them from their scope. One of the few other candidates is RC4, a fast secret-key cryptosystem with variable-length keys.[19] RC4 is adopted in the cellular digital packet data (CDPD) specifications.[20]

Public-key cryptosystems. These algorithms encrypt and decrypt messages with two different keys in such a way that it is difficult to decrypt without the decryption key. The encryption key can be published without compromising security, and is called the public key for this reason; the decryption key is called the private key. Because the encryption and decryption keys in a public-key cryptosystem differ, such systems are often called asymmetric in the literature. The idea comes from Diffie and Hellman.[21]

Public-key cryptosystems provide confidentiality and key management. They can be as secure or more secure than secret-key cryptosystems, but they are generally slower. Their main advantage is that, since the encryption key can be published, parties need not first agree on a secret key. They are often combined with secret-key cryptosystems to gain the benefits of both: speed without prior secrets.

Although there is no primary standard public-key cryptosystem, many consider a cryptosystem invented by Rivest, Shamir, and Adleman (RSA)[22] in 1977 a de facto standard. Public-key cryptosystems, like secret-key cryptosystems, are rarely standardized; when they are standardized, key management is a more likely purpose than confidentiality.

Efforts toward RSA standardization include the intervendor PKCS #1,[23] which gives block formats for RSA operations, and the draft ANSI X9.31 part 4,[24] which is currently based on PKCS #1. PKCS #1's block formats have been adopted by Internet privacy-enhanced mail[25] and, among other algorithms,

Glossary

The acronyms for encryption standards and the groups developing them are considered by some as a form of encryption in its own right. Following is an abridged "key" to the various acronyms and their meanings, as well as to several standards organizations.

ASC X9 Accredited Standards Committee X9 (Financial Services), a body that develops standards for the banking industry; accredited by ANSI

ANSI American National Standards Institute, an organization that accredits standards bodies

CCITT Comité Consultatif International de Télégraphique et Téléphonique, (International Telegraph and Telephone Consultative Committee), an international standards body

CFONB Comité Français d'Organisation et de Normalisation Bancaire, a French banking standards body

DAA Data Authentication Algorithm, a NIST standard authentication code defined in FIPS PUB 113

DEA Data Encryption Algorithm, the secret-key cryptosystem specified by DES

DES Data Encryption Standard, a NIST standard defined in FIPS PUB 46-1 that specifies DEA

Diffie-Hellman A key-agreement algorithm invented by Whitfield Diffie and Martin Hellman

DSA Digital Signature Algorithm, the digital signature scheme specified by DSS

DSS Digital Signature Standard, a proposed NIST standard that specifies DSA

EDE Encrypt-decrypt-encrypt, a mode of DEA involving two keys and three DEA operations that is defined in ANSI X9.17

EES Escrowed Encryption Standard, a proposed NIST standard that specifies Skipjack

FIPS PUB Federal Information Processing Standard publication, one of a series of standards published by NIST

GULS Generic Upper Layers Security, an OSI security architecture effort

IEC International Electrotechnical Commission, an international standards body

IEEE Institute of Electrical and Electronics Engineers, an organization that develops transnational standards; that is, the standards are the consensus of individuals rather than national representatives

Internet A transnational body that develops standards for computer networking and publishes RFCs; also, the network of computers that implements those standards

ISO International Standards Organization, an international standards body

MD2 Message Digest Algorithm 2, a hash function developed by Ron Rivest that is defined in Internet RFC 1319

MD5 Message Digest Algorithm 5, another hash function developed by Ron Rivest and defined in Internet RFC 1321

MDC-2 Manipulation Detection Code 2, the hash function specified in draft ANSI X9.31 part 2

NBS National Bureau of Standards; see NIST

NIST National Institute of Standards and Technology (formerly NBS), a US government agency that develops standards and publishes FIPS PUBs

OIW Open Systems Environment (formerly OSI) Implementors' Workshop, a group of developers that agrees on implementation issues such as algorithms

OSI Open Systems Interconnection, a standard networking model

---

are cited in the OIW implementors' agreements.[26] (As this article was going to press, I received a copy of Australian Standard AS28095.5.3, which specifies RSA.[27])

Digital signature schemes. These schemes "sign" messages and verify the resulting signature with two different keys in such a way that it is difficult to sign without the signing key. Similar to public-key cryptosystems, the verification key can be published without compromising security, and is called the public key; the signing key is called the private key.

Digital signature schemes provide integrity and origin authentication. Like public-key cryptosystems, they do not require that parties first agree on a secret key, and they are generally somewhat slower than, for instance, secret-key cryptosystems and cryptographic hash functions. They are often combined with hash functions to gain the benefits of both.

Public-key cryptosystems and digital signature schemes are

Glossary (continued)

PEM Privacy-enhanced mail, a proposed Internet standard for encrypting and authenticating electronic mail; defined in Internet RFCs 1421-1424

PKCS Public-key cryptography standards, informal standards developed by RSA Laboratories with representatives of Apple, Digital, Lotus, Microsoft, MIT, Northern Telecom, Novell, and Sun; available from RSA Laboratories or via electronic mail to pkcs@rsa.com

RC4 Rivest Cipher 4, a fast secret-key cryptosystem developed by Ron Rivest and proprietary to RSA Data Security

RFC "Request for Comments," an Internet publication

RSA Rivest-Shamir-Adleman algorithm, a public-key cryptosystem and digital signature scheme invented by Ron Rivest, Adi Shamir, and Len Adleman

SC6 Subcommittee 6 (Telecommunications and Information Exchange Between Systems), a joint subcommittee of ISO/IEC

SC27/WG2 Subcommittee 27 (Information Technology), Working Group 2 (Security Techniques), a joint working group of ISO/IEC

SHA Secure Hash Algorithm, the hash function specified by SHS

SHS Secure Hash Standard, a NIST standard defined in FIPS PUB 180 that specifies SHA

SILS Secure Interoperable Local Area Network Security, an IEEE project; also called P802.10

Skipjack The classified secret-key cryptosystem specified by EES

SNMP Simple Network Management Protocol, an Internet standard defined in Internet RFC 1157

Standards

Australia An Australian standards body

X9 See ASC X9

---

closely related. In so-called reversible cryptography, signing in a digital signature scheme is the same as decryption in a public-key cryptosystem, while verification is the same as encryption. In irreversible cryptography, the relationships do not hold, although a given public/private-key pair may work in both a digital signature scheme and a public-key cryptosystem.

There is no primary standard digital signature scheme, but two main efforts are in progress. One involves RSA, which is reversible, and the other involves an irreversible algorithm

proposed by the US National Institute of Standards and Technology (NIST).

ISO/IEC 9796[28] almost creates a standard for RSA, but not quite. It defines a signature block format; RSA is in an informative (but nonstandard) annex. The block format prevents certain mathematical relationships among possible RSA signatures.[29] The draft ANSI X9.31 part 1,[30] which is expected to become a standard late this year, is based on ISO/IEC 9796 and specifies RSA. The intervendor PKCS #1[23] gives alternate block formats for RSA signatures. ISO/IEC's joint working group SC27/WG2 is developing other digital signature standards.

NIST's proposed Digital Signature Standard (DSS),[31] which defines the Digital Signature Algorithm (DSA), has been the center of recent controversy.[32] DSA, an irreversible algorithm, is a variant of signature schemes due to Elgamal[33] and Schnorr.[34] It is intended to be combined with the Secure Hash Algorithm (SHA).[35] Mainly due to objections from industry, DSS has not yet been approved. The draft ANSI X9.30 part 1[36] specifies DSA.

Key-agreement algorithms. These algorithms manage keys through an exchange of messages derived from private values that are not shared. The result of the exchange is that parties agree on a secret key. It is difficult to determine the secret key from the exchanged messages without the private values from which they are derived. Key-agreement algorithms are sometimes called key exchange algorithms in the literature.

Key-agreement algorithms provide confidentiality and key management, and in some cases origin authentication. They do not require that parties first agree on a secret key. As with public-key cryptosystems, no primary standard key-agreement algorithm exists. Many consider an algorithm invented by Diffie and Hellman,[21] usually called Diffie-Hellman, the de facto standard here.

Efforts toward Diffie-Hellman standardization include the intervendor PKCS #3[37] and the draft ANSI X9.30 part 4,[38] which is based on a variant of Diffie-Hellman having origin authentication. The cellular digital packet data (CDPD) specifications[20] adopt Diffie-Hellman key agreement. ISO/IEC's joint working group SC6 is developing standards for key agreement in the network and transport layers of the OSI reference model,[39,40] with Diffie-Hellman as a possible algorithm.

Cryptographic hash functions. These functions reduce a message of arbitrary length to a short code so that it is difficult to find a message with a given hash code, and in some cases also to find two messages with the same hash code. There is no key. Hash functions are also called message digests and modification detection codes in the literature.

A hash code is typically 128 or 160 bits long. Ideally, an attacker's only approach is trial and error, which amounts to 2^{128} trials to find a message with a given hash code (for a 128-bit hash), and 2^{64} trials to find two messages with the same hash code. (This is akin to the "birthday paradox": You need

365 people in a room to be likely to find one with a given birthday, but only 23 to be likely to find two with the same birthday.) Hash functions are generally quite fast. They provide message integrity to parties knowing a message's hash code. They are often combined with digital signature schemes, as noted earlier.

The Secure Hash Standard (SHS),[35] which defines SHA, is the primary standard. SHA produces a 160-bit hash from a message of arbitrary length; it is intended to be combined with DSA.[31] ANSI X9.30 part 2[41] specifies SHA.

Other hash algorithms suitable for standardization include MD2 and MD5, developed by Ron Rivest for RSA Data Security[12,42] and adopted by Internet privacy-enhanced mail,[25] and MDC-2, which is specified in draft ANSI X9.31 part 2.[44] SC27/WG2 is also developing standards for hash functions.

Authentication codes. These codes reduce a message of arbitrary length to a short code under a secret key so that it is difficult, without the key, to compute the authentication code, or to find a new message with a given authentication code. Authentication codes provide message integrity and origin authentication to parties who have previously agreed on a secret key. The message itself need not be encrypted.

An authentication code is typically 32 or 64 bits long, and the keys are 56 bits long. Ideally, an attacker's only approach is trial and error on the keys; arbitrary message modifications have some probability of success, but the attacker cannot check for success without the help of the real user. Authentication codes, like hash functions, are generally quite fast.

The primary standard is FIPS PUB 113,[45] which defines the Data Authentication Algorithm. The algorithm is a variant of DEA; it produces a 32-bit authentication code from a message of arbitrary length and a 56-bit key. ANSI X9.9[46] and Australian standard AS2805.4[47] specify DAA.

Applications

The applications standards described next combine families of algorithms, and sometimes specify particular algorithms, to solve confidentiality, integrity, origin authentication, and key management problems. Although many of the standards specify much more than just cryptography, encryption plays an important role.

Ideally, an algorithm should work in many applications, and many algorithms should work in a given application. The design of applications and algorithms is in this sense "orthogonal," and the designers have generally done a good job at providing orthogonality.

Do not confuse these applications with the applications layer of the OSI reference model; some may well run at that layer, and others at lower layers.

Secure electronic mail. Six years in development and now a proposed standard, Internet privacy-enhanced mail (PEM) combines secret-key cryptosystems, public-key cryptosystems, hash functions, and digital signature schemes

to provide security for electronic mail.[48] It is a text-based protocol compatible with most electronic-mail systems. PEM supports public-key and secret-key techniques; the former involves X.509 certificates.[49] Currently, PEM has adopted RSA, DEA, MD2, and MD5 algorithms,[25] but the protocols are flexible and other suites of algorithms are likely to be added.

Mail is not the only application of PEM, of course, although it is a primary one. The same protocol that adds encryption or authentication to a mail message can enhance any digital document, such as a contract; the document need not be mailed to someone.

The intervendor PKCS #7[50] is a binary extension of PEM; it offers the same services, but works with binary data and allows one to sign attributes such as the time of day along with the underlying message. Certain modes of PKCS #7 are cryptographically compatible with PEM, in the sense that messages can be translated between the two protocols without any cryptographic operations. PKCS #7 does not specify a particular algorithm.

Another approach to electronic-mail security is found in X.400 message-handling systems,[51] which solve the basic problems of confidentiality, authentication, and key management. X.400 also provides special encryption-based services such as proof of submission and proof of delivery. (X.411 supplies the details.[52]) X.400, like most international standards, does not specify particular algorithms. It supports both public-key and secret-key techniques. ISO 10021-1[53] is technically aligned with X.400.

X.435,[54] a standard for electronic data interchange over X.400, builds on X.411's services, defining related services such as signed receipts.

Secure communications. These standards focus on the security of local-area networks and wireless links. IEEE's P802.10 project, Secure Interoperable LAN (local area network) Security (SILS), addresses privacy and authentication of data at the data link layer. Devices following the protocol encrypt data link frames as they pass through the network; the protocol is transparent to higher layers. A proposed draft[55] specifies Diffie-Hellman key agreement. The CDPD specifications[20] define an encryption protocol for wireless links based on Diffie-Hellman key agreement and RC4.

IEEE project P802.11, focusing on wireless links, has just started.

Directory authentication and network management. X.509 directory authentication[49] applies public-key and secret-key techniques to the problem of determining the identity of a user attempting to access an X.500 global directory.[56] "Weak" authentication identifies a user by a password, while "strong" authentication involves digital signatures. The authentication protocols can also ensure that messages to and from the directory are not modified in transit.

X.509 standardizes on no particular algorithm, although RSA is in an informative annex. Two additional contributions

of X.509 are certificates, which bind a public key to a user's name with a digital signature, and certificate-revocation lists, which break the binding. These elements have found their way into other applications such as privacy-enhanced mail and the X9.30 and X9.31 drafts. Although directories are just emerging, users' names in the related applications are designed in anticipation of a future directory entry. ISO 9594-8[57] is technically aligned with X.509.

In a proposed security standard for the Internet's Simple Network Management Protocol (SNMP),[58] parties identify each other with a secret shared key.[59] Network management requests are hashed together with the secret key under MD5 to produce an authentication code. Encryption with DEA is also an option.

SC27/WG2 is developing authentication protocols involving public-key and secret-key techniques.

Banking. The primary key management standard for the banking industry is ANSI X9.17. It is based entirely on DEA and related algorithms, including the EDE mode of DEA. To date, X9's standards have all involved secret-key techniques; work on public-key techniques is in progress in X9.30 and X9.31. Other banking standards efforts include

- draft Australian standard AS2805.6.5.3,[60] which specifies RSA;
- CFONB ETEBAC-5,[61] a French banking standard that specifies RSA and DEA; and
- ISO CD 11666, a draft standard for banking key management that specifies RSA.[62,63] Whether it will be approved is unclear, as its architectural features have been criticized.[64]

Escrowed encryption. A likely candidate to surpass even the DSS controversy is the proposed Escrowed Encryption Standard (EES),[17] part of the US government's Capstone project for encryption standards. It implements an April 1993 presidential order that certain encryption devices provide entry points for legitimate law-enforcement wiretaps. The government's Clipper chips are the first examples of such devices.[65]

EES is based on the Skipjack algorithm and involves a classified law-enforcement access field (LEAF). Each hardware device complying with EES (software is not allowed) has a secret key; the key is split at the factory and "escrowed" with (that is, put into the custody of, as with money or deeds) two government agencies. Under court order, the agencies reconstruct the key. With the secret key and LEAF, authorized officials can decrypt messages encrypted by the device. Neither escrow agency can decrypt messages by itself.

What is controversial about EES appears not so much to be government wiretapping, which has always been controversial, but the issues of algorithm secrecy, hardware-only implementation, and potential security risks in the manufac-

turing and key escrow processes. The panel that reviewed the Skipjack algorithm is also evaluating the manufacturing and key escrow processes.

CRYPTOGRAPHY IS FINDING BROAD APPLICATION in the computer world. There is much common ground in the underlying algorithms. Interestingly, solutions to the confidentiality problem—encryption in the pure sense—seem to be the hardest to standardize. Much more activity focuses on peripheral cryptographic problems such as authentication and key management, as well as algorithm-independent standards.

As evidenced by the parallel X9.30 and X9.31 efforts, the controversy over DSS has brought about parallel standards, one involving the reversible model (for example, RSA). Here, signing is the same as encryption, and verification is the same as decryption. The other standard involves the irreversible model (for example, DSA) without such relationships. Reversibility is considered by some to open the door to confidentiality of unlimited security, a problematic feature for law enforcement and national security concerns. Others see dual standardization to be problematic for industry concerns.

Since NIST may have reaffirmed DES for the last time, what comes next? The Internet's PEM working group has been looking at new encryption algorithms, among them the so-called triple-DES with three DEA operations, of which X9.17's EDE is one example. Whether the factor-of-three slowdown in performance is too much remains to be seen, but in light of the secrecy around the Skipjack algorithm and the few published alternatives, most likely triple-DES will become a standard encryption algorithm in some corner of the standards world. RC4 may play a role as well.

While all of this is sorting itself out, a new IEEE project, sponsored by the Computer Society's Microprocessor and Microcomputer Standards Committee, aims to complete the family of public-key standards. These standards will be based on the RSA and Diffie-Hellman algorithms, covering key management, encryption, authentication, key generation, and hardware support. The IEEE authorized P1363, "RSA, Diffie-Hellman, and related public-key techniques" this June, and an initial meeting is being planned as of this writing. ∎

References

1. W. Diffie, "The First Ten Years of Public-Key Cryptography," *Proc. IEEE*, 1988, pp. 560-577.
2. G.J. Simmons, ed., *Contemporary Cryptology: The Science of*

Information Integrity, IEEE, New York, 1992.

3. P. Fahn, *Answers to Frequently Asked Questions About Today's Cryptography, Version 2.0*, RSA Laboratories, Redwood City, Calif., Sept. 1993.

4. A. Patel, "Emerging Network Security Standards in an OSI Environment," *Computer Standards & Interfaces*, 1989/1990, pp. 239-247.

5. *Recommendation X.800: Security Architecture for Open Systems Interconnection for CCITT Applications*, CCITT, Geneva, 1991.

6. *ISO/IEC DIS 11586: Generic Upper Layers Security*, ISO/IEC, Geneva, 1993.

7. Accredited Standards Committee X9, *Working Draft: American National Standard X9.41-1993: Security Services Management for the Financial Services Industry*, American Bankers Assoc., Washington, D.C., Aug. 1993.

8. *DoD 5200.28-STD: Department of Defense (DoD) Trusted Computer System Evaluation Criteria (TCSEC)*, US Department of Defense, Washington, D.C., 1985.

9. *FIPS Publication 46-1: Data Encryption Standard*, NIST, Washington, D.C., Jan. 22, 1988; originally issued by the National Bureau of Standards.

10. Accredited Standards Committee X3, *ANSI X3.92: Data Encryption Algorithm (DEA)*, ANSI, New York, 1981.

11. *Australian Standard 2805.5 1985: Electronics Funds Transfer—Requirements for Interfaces: Part 5—Data Encryption Algorithm*, Standards Assoc. of Australia, North Sydney, NSW, 1985.

12. E. Biham and A. Shamir, "Differential Cryptanalysis of the Full 16-Round DES," *Proc. Crypto 92, Advances in Cryptology*, Springer-Verlag, New York, 1993, to appear.

13. D. Coppersmith, "The Data Encryption Standard (DES) and Its Strength Against Attacks," tech. report RC 18613 (81421), IBM Research Div., Yorktown Heights, N.Y., Dec. 1992.

14. *FIPS Publication 81: DES Modes of Operation*, NIST, Dec. 2, 1980.

15. Accredited Standards Committee X9, *American National Standard X9.17: Financial Institution Key Management (Wholesale)*, ANSI, 1985.

16. *PKCS #5: Password-Based Encryption Standard, Version 1.4*, RSA Data Security, Inc., Redwood City, Calif., June 1991.

17. NIST, "A Proposed Federal Information Processing Standard for an Escrowed Encryption Standard (EES)," *Federal Register*, Vol. 58, No. 145, July 30, 1993.

18. E.F. Brickell et al., "Skipjack Review, Interim Report: The Skipjack Algorithm," July 28, 1993; contact author for copies.

19. R.L. Rivest, *The RC4 Encryption Algorithm*, RSA Data Security, Inc., Mar. 12, 1992.

20. Ameritech Mobile Communications et al., *Cellular Digital Packet Data System Specifications: Part 406: Airlink Security*, CDPD Industry Input Coordinator, Costa Mesa, Calif., July 1993.

21. W. Diffie and M.E. Hellman, "New Directions in Cryptography," *IEEE Trans. Information Theory*, Vol. IT-22, 1976, pp. 644-654.

22. R.L. Rivest, A. Shamir, and L. Adleman, "A Method for Obtaining Digital Signatures and Public-Key Cryptosystems," *Comm. ACM*, Vol. 21, No. 2, Feb. 1978, pp. 120-126.

23. *PKCS #1: RSA Encryption Standard, Version 1.4*, RSA Data Security, Inc., June 1991.

24. Accredited Standards Committee X9, *Working Draft: American National Standard X9.31-1993: Public Key Cryptography Using Reversible Algorithms for the Financial Services Industry: Part 4: Management of Symmetric Algorithm Keys Using RSA*, Am. Bankers Assoc., June 4, 1993.

25. D. Balenson, *RFC 1423: Privacy Enhancement for Internet Electronic Mail: Part III—Algorithms, Modes, and Identifiers*, Trusted Information Systems, Inc., Glenwood, Md., Feb. 1993.

26. *Special Publication 500-183: Stable Implementation Agreements for Open Systems Interconnection Protocols: Part 12—OS Security*, NIST, June 1992.

27. *Australian Standard 2805.5.3—Electronic Data Transfer—Requirements for Interfaces: Part 5.3: Data Encipherment Algorithm 2*, Standards Assoc. of Australia, 1992.

28. *International Standard 9796: Information Technology, Security Techniques: Digital Signature Scheme Giving Message Recovery*, ISO/IEC, 1991.

29. L.C. Guillou et al., "Precautions Taken Against Various Potential Attacks in ISO/IEC DIS 9796," *Proc. Eurocrypt 90, Advances in Cryptology*, I.B. Damgard, ed., Springer-Verlag, 1991, pp. 465-473.

30. Accredited Standards Committee X9, *Working Draft: American National Standard X9.31-1992: Public Key Cryptography Using Reversible Algorithms for the Financial Services Industry: Part 1: The RSA Signature Algorithm*, Am. Bankers Assoc., Mar. 7, 1993.

31. *Publication XX: Announcement and Specifications for a Digital Signature Standard (DSS)*, NIST, Aug. 19, 1992.

32. NIST, "The Digital Signature Standard, Proposal and Discussion," *Comm. ACM*, Vol. 35, No. 7, July 1992, pp. 36-54.

33. T. Elgamal, "A Public-Key Cryptosystem and a Signature Scheme Based on Discrete Logarithms," *IEEE Trans. Information Theory*, Vol. IT-31, 1985, pp. 469-472.

34. C.P. Schnorr, "Efficient Identification and Signatures for Smart Cards," *Proc. Crypto 89, Advances in Cryptology*, Springer-Verlag, 1990, pp. 239-251.

35. *FIPS Publication 180: Secure Hash Standard (SHS)*, NIST, May 11, 1993.

36. Accredited Standards Committee X9, *Working Draft: American National Standard X9.30-199X: Public Key Cryptography Using Irreversible Algorithms for the Financial Services Industry: Part 1: The Digital Signature Algorithm (DSA)*, Am. Bankers Assoc., June 18, 1993.

37. *PKCS #3: Diffie-Hellman Key Agreement Standard, Version 1.3*, RSA Data Security, Inc., June 1991.

38. Accredited Standards Committee X9, *Working Draft: American National Standard X9.30-1993: Public Key Cryptography Using Irreversible Algorithms for the Financial Services Industry: Part 4: Management of Symmetric Algorithm Keys Using Diffie-Hellman*, Am. Bankers Assoc., June 4, 1993.

39. ISO/IEC JTC1/SC6, *N2559: Draft Network Layer Security Protocol*, ISO/IEC, Sept. 1990.

40. ISO/IEC JTC1/SC6, *N6285: Draft Transport Layer Security Protocol*, ISO/IEC, Nov. 1990.

41. Accredited Standards Committee X9, *Working Draft: American National Standard X9.30-1993: Public Key Cryptography Using Irreversible Algorithms for the Financial Services Industry: Part 2: The Secure Hash Algorithm (SHA)*, Am. Bankers Assoc., 1993.

42. B. Kaliski, *RFC 1319: The MD2 Message-Digest Algorithm*, RSA Data Security, Inc., Apr. 1992.

43. R.L. Rivest, *RFC 1321: The MD5 Message-Digest Algorithm*, MIT Laboratory for Computer Science, Cambridge, Mass., and RSA Data Security, Inc., Apr. 1992.

44. Accredited Standards Committee X9, *Working Draft: American National Standard X9.31-1992: Public Key Cryptography Using Reversible Algorithms for the Financial Services Industry: Part 2: The MDC-2 Hash Algorithm*, Am. Bankers Assoc., June 4, 1993.

45. *FIPS Publication 113: Computer Data Authentication*, NIST, May 30, 1985.

46. Accredited Standards Committee X9, *American National Standard X9.9: Financial Institution Message Authentication*, ANSI, 1982.

47. *Australian Standard 2805.4 1985: Electronics Funds Transfer—Requirements for Interfaces: Part 4—Message Authentication*, Standards Assoc. of Australia, 1985.

48. J. Linn, *RFC 1421: Privacy Enhancement for Internet Electronic Mail: Part I—Message Encipherment and Authentication Procedures*, Internet Activities Board, Feb. 1993.

49. *Recommendation X.509: The Directory—Authentication Framework*, CCITT, 1988.

50. *PKCS #7: Cryptographic Message Syntax Standard, Version 1.4*, RSA Data Security, Inc., June 1991.

51. *Recommendation X.400: Message Handling System and Service Overview*, CCITT, 1988.

52. *Recommendation X.411: Message Handling Systems: Message Transfer System: Abstract Service Definition and Procedures*, CCITT, 1988.

53. *International Standard 10021-1: Information Processing Systems, Text Communication, MOTIS, Message Handling: System and Service Overview*, ISO/IEC, 1988.

54. *Recommendation X.435: Message Handling Systems: EDI Messaging System*, CCITT, 1991.

55. *IEEE Draft Standard for Interoperable Local Area Network (LAN) Security: Part C—Key Management Proposal*, IEEE, Apr. 10, 1992.

56. *Recommendation X.500: The Directory—Overview of Concepts, Models and Services*, CCITT, 1988.

57. *International Std. 9594-8: Information Processing Systems—Open Systems Interconnection—The Directory—Authentication Framework*, ISO/IEC, 1988.

58. J. Case et al., *RFC 1157: The Simple Network Management Protocol*, SNMP Research, Knoxville, Tenn., May 1990.

59. J. Galvin, K. McCloghrie, and J. Davin, *RFC 1352: SNMP Security Protocols*, Trusted Information Systems, Inc., July 1992.

60. *Draft Australian Standard: Electronics Funds Transfer—Requirements for Interfaces: Part 6.5.3—Key Management—Terminal Initialisation—Asymmetrical*, Standards Assoc. of Australia, May 1990.

61. *Echanges Télématiques entre les Banques et leurs Clients, Standard ETEBAC 5, v1.1*, Comité Français d'Organisation et de Normalisation Bancaire, Paris, 1989.

62. *ISO CD 11166-1, Banking—Key Management by Means of Asymmetric Algorithms—Part 1: Principles, Procedures and Formats*, ISO TC68/SC2/N346, 1992.

63. *ISO CD 11166-2, Banking—Key Management by Means of Asymmetric Algorithms—Part 2: Approved Algorithms Using the RSA Cryptosystem*, ISO TC68/SC2/N304, 1992.

64. ISO/IEC JTC1/SC27/WG2, *N205: Criticism of ISO CD 11666 Banking—Key Management by Means of Asymmetric Algorithms*, ISO/IEC, Apr. 1993.

65. D.E. Denning, "The Clipper Encryption System," *Am. Scientist*, Vol. 81, No. 4, July-Aug.1993, pp. 319-323.

Burt Kaliski is chief scientist of the RSA Laboratories division of RSA Data Security. His research interests include cryptography and fast arithmetic.

Kaliski received the BS, MS, and PhD degrees in computer science from the Massachusetts Institute of Technology, Cambridge. He is a member of the IEEE Computer Society, the International Association for Cryptologic Research, Sigma Xi, and Tau Beta Pi. He served as general chair of Crypto 91, and he is interim chair of IEEE P1363.

Direct questions concerning this article to the author at RSA Laboratories, 100 Marine Parkway, Redwood City, CA 94065; or burt@rsa.com.

Reader Interest Survey

Indicate your interest in this article by circling the appropriate information on the Reader Service Card.

Low 171   Medium 172   High 173

1993 Index
IEEE MICRO
Vol. 10

Mateosian, Richard. Review of 'The Computer Curmudgeon' (Kawasaki, G.; 1992); *M-M Apr 93* 81

Mateosian, Richard. Review of 'The Computer Font Book' (Searfoss, G.; 1993); *M-M Apr 93* 82

Mateosian, Richard. Review of 'The Windows 3.1 Font Book' (Angell, D., and Heslop, B.; 1992); *M-M Apr 93* 82

Mateosian, Richard. Review of 'High-Speed Digital Design—A Handbook of Black Magic' (Johnson, H. W., and Graham, M.; 1993); *M-M Jun 93* 85-86

Mateosian, Richard. Review of 'Intel's SL Architecture—Designing for Portable Applications' (Yuen, D.; 1993); *M-M Jun 93* 86

Mateosian, Richard. Review of 'Code Complete' (McConnell, S.; 1993); *M-M Aug 93* 61-62

Mateosian, Richard. Review of 'Programming As If People Mattered' (Broenstein, N. S.; 1991); *M-M Aug 93* 61

Mateosian, Richard. Review of 'Writing Solid Code' (Maguire, S.; 1993); *M-M Aug 93* 62

Mateosian, Richard. Micro Review of 'Inside Visual C++' (Kruglinski, D. J.; 1993); *M-M Oct 93* 5

Mateosian, Richard. Review of 'The Underground Guide to Laser Printers' (*Flash Magazine* editors; 1993); *M-M Dec 93* 92-93

Mateosian, Richard. Review of 'The Windows Programming Puzzle Book' (Crouse, K.; 1993); *M-M Dec 93* 93

Mateosian, Richard. Review of 'The Visual Guide to Visual Basic for Windows,' 2nd ed. (Mansfield, R.; 1993); *M-M Dec 93* 93-94

Matsui, Shigezumi, *see* Uchiyama, Kunio, *M-M Oct 93* 12-22

McLellan, Edward. The Alpha AXP architecture and 21064 processor; *M-M Jun 93* 36-47

McWilliams, G. Jeanette, *see* White, Steven W., *M-M Aug 93* 8-16

McWilliams, G. Jeanette, *see* White, Steven W., *M-M Oct 93* 79-89

Miller, Larry. On the Edge—Desktop document management; *M-M Jun 93* 4-6

Miller, Patrick M., *see* Hurson, A. R., *M-M Apr 93* 59-65

Misunas, David, *Guest Ed.* Advanced packaging and interconnection technology (special issue intro.); *M-M Apr 93* 7-9

Moore, Charles R., *see* Becker, Michael C., *M-M Oct 93* 54-68

Muhich, John S., *see* Becker, Michael C., *M-M Oct 93* 54-68

Myers, Ware. Micro View—ICs per vehicle increasing rapidly; *M-M Feb 93* 4-6

Myers, Ware. Micro View—Market volume drives neural-net technology; *M-M Apr 93* 3-5

N

Nakagawa, Norio, *see* Uchiyama, Kunio, *M-M Oct 93* 12-22

Nakamura, Kazuo, Narumi Sakashita, Yasuhiko Nitta, Ken'ichi Shimomura, and Takeshi Tokuda. Fuzzy inference and fuzzy inference processor; *M-M Oct 93* 37-48

Narita, Susumu, *see* Uchiyama, Kunio, *M-M Oct 93* 12-22

Neusser, Stefan, Jos Nijhuis, Lambert Spaanenburg, Bernd Hoefflinger, Uwe Franke, and Hans Fritz. Neurocontrol for lateral vehicle guidance; *M-M Feb 93* 57-66

Nijhuis, Jos, *see* Neusser, Stefan, *M-M Feb 93* 57-66

Nitta, Yasuhiko, *see* Nakamura, Kazuo, *M-M Oct 93* 37-48

Noble, John F. Review of 'Softwars: The Legal Battle for Control of the Global Software Industry' (Clapes, A. L.); *M-M Jun 93* 78-80

P

Palmquist, Ulf. Intelligent cruise control and roadside information; *M-M Feb 93* 20-28

Parkin, Mike, *see* Agarwal, Anant, *M-M Jun 93* 48-61

Pavan, Paolo, *see* Zanoni, Enrico, *M-M Feb 93* 30-48

Pnevmatikatos, Dionisios N., *see* Gee, Jeffrey D., *M-M Aug 93* 17-27

Price, Dick. Micro View—Berkeley's Patterson predicts exciting times; *M-M Aug 93* 4-5, 79

R

Russell, Gordon, and Paul Shaw. Shifting register windows; *M-M Aug 93* 28-35

S

Sakamura, Ken, *Guest Ed.* Toward a world filled with computers (intro. to special issue on three powerful new chips from Japan); *M-M Oct 93* 6-11

Sakashita, Narumi, *see* Nakamura, Kazuo, *M-M Oct 93* 37-48

Seger, Ulrich, Heinz-Gerd Graf, and Marc E. Landgraf. Vision assistance in scenes with extreme contrast; *M-M Feb 93* 50-56

Shaw, Paul, *see* Russell, Gordon, *M-M Aug 93* 28-35

Sherwani, Naveed A., *see* Burman, Surendra, *M-M Apr 93* 28-35

Shimomura, Ken'ichi, *see* Nakamura, Kazuo, *M-M Oct 93* 37-48

Shladover, Steven E. Research and development needs for advanced vehicle control systems; *M-M Feb 93* 11-19

Smith, Alan Jay, *see* Gee, Jeffrey D., *M-M Aug 93* 17-27

Some, Rafael, Myles Suer, and John Carson. On the edge—The development of 3D memory for high and low-end systems; *M-M Dec 93* 6-7, 102

Spaanenburg, Lambert, *see* Neusser, Stefan, *M-M Feb 93* 57-66

Stern, Richard H. Micro Law—A guardedly cheerful note—For a change; *M-M Feb 93* 73-75

Stern, Richard H. Micro Law—A back door to protecting look and feel?; *M-M Apr 93* 74-76

Stern, Richard H. Micro Law—Glitches left in software copyright system; *M-M Jun 93* 75-77

Stern, Richard H. Micro Law—Reverse engineering for future compatibility; *M-M Aug 93* 6-7, 76-79

Stern, Richard H. Micro Law—Protecting industrial property rights; *M-M Oct 93* 2-3, 100-103

Stern, Richard H. Micro Law—Patents and international trade issues; *M-M Dec 93* 89-92

Suer, Myles, *see* Some, Rafael, *M-M Dec 93* 6-7,102

T

Takahashi, Hiromasa, *see* Awaga, Makoto, *M-M Oct 93* 24-36

Talia, Domenico. Message-routing systems for transputer-based multicomputers; *M-M Jun 93* 62-72

Tokuda, Takeshi, *see* Nakamura, Kazuo, *M-M Oct 93* 37-48

Tuttle, David P., *see* Becker, Michael C., *M-M Oct 93* 54-68

U

Uchiyama, Kunio, Fumio Arakawa, Susumu Narita, Hirokazu Aoki, Ikuya Kawasaki, Shigezumi Matsui, Mitsuyoshi Yamamoto, Norio Nakagawa, and Ikuo Kudo. The Gmicro/500 superscalar microprocessor with branch buffers; *M-M Oct 93* 12-22

Updegrove, Andrew. Forming, funding and operating standard-setting consortia; *M-M Dec 93* 52-61

V

Vickers, Richard. The development of ATM standards and technology—A retrospective view; *M-M Dec 93* 62-73

W

Wang, Chia-Jiu, and Frank Emnett. Implementing precise interruptions in pipelined RISC processors; *M-M Aug 93* 36-43

Weiner, Bill, *see* Asprey, Tom, *M-M Jun 93* 22-35

White, Steven W., Phil D. Hester, Jack W. Kemp, and G. Jeanette McWilliams. How does processor MHz relate to end-user performance?—1: Pipelines and functional units; *M-M Aug 93* 8-16

White, Steven W., Phil D. Hester, Jack W. Kemp, and G. Jeanette McWilliams. How does processor MHz relate to end-user performance?—2: Memory subsystem and instruction set; *M-M Oct 93* 79-89

Y

Yamamoto, Mitsuyoshi, *see* Uchiyama, Kunio, *M-M Oct 93* 12-22

Yetter, Jeff, *see* Asprey, Tom, *M-M Jun 93* 22-35

Yeung, Donald, *see* Agarwal, Anant, *M-M Jun 93* 48-61

Z

Zanoni, Enrico, and Paolo Pavan. Improving the reliability and safety of automotive electronics; *M-M Feb 93* 30-48

SUBJECT INDEX
A

Application-specific integrated circuits

cooperation between private and public sectors and foreign entities in several Japanese high-tech industries (Software Report). *Kahaner, David K., M-M Oct 93* 90-92

programmable multichip module design principles and applications. *Burman, Surendra, + , M-M Apr 93* 28-35

Arithmetic; cf. Coprocessors; Floating-point arithmetic

Assembly systems

implementation of fully automated assembly lines in Japan (Software Report). *Kahaner, David K., M-M Jun 93* 88-92

VLSI silicon resources perspective on implementing precise interruptions in pipelined RISC processors. *Wang, Chia-Jiu, +, M-M Aug 93* 36-43

Computer interfaces
 book review; The Windows 3.1 Font Book (Angell, D., and Heslop, B.; 1992). *Mateosian, Richard, M-M Apr 93* 82
 development of virtual reality systems in Japan; application to 3-D imaging of software structures and remote control of construction robots. *Kahaner, David K., M-M Apr 93* 66-73

Computer interfaces, human factors
 book review; Programming As If People Mattered (Broenstein, N. S.; 1991). *Mateosian, Richard, M-M Aug 93* 61
 book review; The Elements of Friendly Software Design—The New Edition (Heckel, P; 1991). *Mateosian, Richard, M-M Feb 93* 79

Computer language processors; cf. Compilers

Computer languages; cf. Visual languages

Computer networks
 economic perspective on computer network development and standardization management. *Greenstein, Shane, M-M Dec 93* 36-51

Computer networks; cf. Multiprocessing, interconnection

Computer operating systems; cf. Software, operating systems

Computer performance
 cache performance of SPEC92 benchmark suite. *Gee, Jeffrey D., +, M-M Aug 93* 17-27
 IBM RISC system/6000 Model 580 and DEC Alpha processors; comparison of pipelines and interaction between functional units. *White, Steven W., +, M-M Aug 93* 8-16
 IBM RISC System/6000 Model 80 and DEC Alpha processors; comparison of memory effects and instruction set differences . *White, Steven W., +, M-M Oct 93* 79-89
 shifting register windows to reduce spill/refill memory overhead. *Russell, Gordon, +, M-M Aug 93* 28-35
 Spearmints system of hardware components supporting performance measurements in distributed systems. *Kleinhans, Uwe, +, M-M Oct 93* 69-78

Computer peripherals; cf. Computer interfaces; Disk drives

Computer pipeline processing; cf. Pipeline processing

Computers; cf. Distributed computing; Multiprocessing; Parallel processing

Content-addressable memories; cf. Associative memories

Control systems; cf. Road vehicle control

Coprocessors
 μVP 64-b vector coprocessor for 206-MFLOPS numerical computation. *Awaga, Makoto, +, M-M Oct 93* 24-36

Copyright protection
 conflict between intellectual property laws and concepts of standards, and compatibility in the computer software and hardware industries. *Davis, G. Gervaise, III, M-M Dec 93* 19-27

Copyright protection; cf. Software protection

Cryptography
 survey of encryption standards for computer networks. *Kaliski, Burt, M-M Dec 93* 74-81

D

Data communication; cf. Computer networks; Integrated services digital networks; Multiprocessing, interconnection

Data management; cf. Memory management

Data security; cf. Cryptography

Desktop publishing; book review; The Underground Guide to Laser Printers (*Flash Magazine* eds.; 1993). *Mateosian, Richard M., M-M Dec 93* 92-93

Digital communication; cf. Integrated services digital networks

Digital image processing; cf. Image processing

Digital signal processors; cf. Signal processing

Disk drives
 cooperation between private and public sectors and foreign entities in several Japanese high-tech industries (Software Report). *Kahaner, David K., M-M Oct 93* 90-92

Distributed computing
 Spearmints system of hardware components supporting performance measurements in distributed systems. *Kleinhans, Uwe, +, M-M Oct 93* 69-78

Document handling
 technical requirements of desktop document management (On the Edge). *Miller, Larry, M-M Jun 93* 4-6

E

Economics; cf. Business economics; Computer economics

F

Fabrication; cf. Thin-film circuit fabrication

Feedforward systems
 feedforward neural networks for lateral road vehicle guidance. *Neusser, Stefan, +, M-M Feb 93* 57-66

FET integrated circuits; cf. CMOS integrated circuits

Floating-point arithmetic
 architecture of Pentium microprocessor. *Alpert, Donald, +, M-M Jun 93* 11-21
 performance features of PA7100 microprocessor. *Asprey, Tom, +, M-M Jun 93* 22-35
 μVP 64-b vector coprocessor for 206-MFLOPS numerical computation. *Awaga, Makoto, +, M-M Oct 93* 24-36

Fuzzy control
 theory of fuzzy inference; application in inference processor for fuzzy control. *Nakamura, Kazuo, +, M-M Oct 93* 37-48

G

Gallium FETs; cf. MESFET integrated circuits

H

HDTV
 HDTV research in Japan. *Kahaner, David K., M-M Oct 93* 49-53

High-definition TV; cf. HDTV

High-speed circuits/devices
 book review; High-Speed Digital Design—A Handbook of Black Magic (Johnson, H. W., and Graham, M.; 1993). *Mateosian, Richard, M-M Jun 93* 85-86

High-speed integrated circuits
 Gmicro/500 superscalar microprocessor with branch buffers and dual-pipeline structure for high-speed instruction execution. *Uchiyama, Kunio, +, M-M Oct 93* 12-22

Human factors; cf. Computer interfaces, human factors

Hybrid integrated circuit fabrication; cf. Thin-film circuit fabrication

Hybrid integrated circuit interconnections; cf. Thin-film circuit interconnections

Hybrid integrated circuits; cf. Surface mounting

I

Iconic languages; cf. Visual languages

IEEE Computer Society; cf. Awards

IEEE standards
 process of IEEE standards development (Micro Standards). *Diamond, Stephen L., M-M Aug 93* 59-60

Image processing
 development of virtual reality systems in Japan; application to 3-D imaging of software structures and remote control of construction robots. *Kahaner, David K., M-M Apr 93* 66-73
 performance and architecture of high-dynamic-range camera chip for vision systems in traffic environments. *Seger, Ulrich, +, M-M Feb 93* 50-56

Image sensors
 performance and architecture of high-dynamic-range camera chip for vision systems in traffic environments. *Seger, Ulrich, +, M-M Feb 93* 50-56

India
 trends in Chinese and Indian computer industries; establishment of UN International Institute for Software Technology in China. *Kahaner, David K., M-M Aug 93* 44-48

Industrial control; cf. Assembly systems

Information retrieval; cf. Document handling

Innovation; cf. Technological innovation

Integrated circuit bonding
 amalgams for improved electronics interconnection. *MacKay, Colin A., M-M Apr 93* 46-58

Integrated circuit bonding; cf. Surface mounting

Integrated circuit design
 design of 16-kb Θ-search associative memory. *Hurson, A. R., +, M-M Apr 93* 59-65

Integrated circuit fabrication; cf. Integrated circuit bonding; Integrated circuit packaging; Thin-film circuit fabrication

Integrated circuit interconnections
 advanced packaging and interconnection technology (special issue). *M-M Apr 93* 7-44, 46-58

N

Networks; cf. Computer networks; Multiprocessing, interconnection; Neural networks

Neural networks

feedforward neural networks for lateral road vehicle guidance. *Neusser, Stefan, + , M-M Feb 93* 57-66

interview with F. Faggin on neural network technology (Micro View). *Myers, Ware, M-M Apr 93* 3-5

O

Office automation; cf. Document handling

Open Systems Interconnection

meaning of open systems and standards (Micro Standards). *Diamond, Stephen L., M-M Feb 93* 71-72

Operating systems; cf. Software, operating systems

OSI; cf. Open Systems Interconnection

P

Packaging; cf. Integrated circuit packaging

Packet switching; cf. Asynchronous transfer mode

Parallel processing

trends in Chinese and Indian computer industries; establishment of UN International Institute for Software Technology in China. *Kahaner, David K., M-M Aug 93* 44-48

Parallel processing; cf. Multiprocessing; Pipeline processing

Parallel processing, interconnection; cf. Multiprocessing, interconnection

Pipeline processing

Alpha AXP architecture supporting multiprocessing; implementation in 21064 processor supporting multiprocessing. *McLellan, Edward, M-M Jun 93* 36-47

architecture of Pentium microprocessor. *Alpert, Donald, + , M-M Jun 93* 11-21

Gmicro/500 superscalar microprocessor with branch buffers and dual-pipeline structure for high-speed instruction execution. *Uchiyama, Kunio, + , M-M Oct 93* 12-22

IBM RISC system/6000 Model 580 and DEC Alpha processors; comparison of pipelines and interaction between functional units. *White, Steven W., + , M-M Aug 93* 8-16

IBM RISC System/6000 Model 80 and DEC Alpha processors; comparison of memory effects and instruction set differences . *White, Steven W., + , M-M Oct 93* 79-89

performance features of PA7100 microprocessor. *Asprey, Tom, + , M-M Jun 93* 22-35

VLSI silicon resources perspective on implementing precise interruptions in pipelined RISC processors. *Wang, Chia-Jiu, + , M-M Aug 93* 36-43

Pipeline processing; cf. Vector processing

Printed circuits; cf. Surface mounting

Printing

book review; The Computer Font Book (Searfoss, G.; 1993). *Mateosian, Richard, M-M Apr 93* 82

book review; The Windows 3.1 Font Book (Angell, D., and Heslop, B.; 1992). *Mateosian, Richard, M-M Apr 93* 82

Privacy; cf. Cryptography

Programmable circuits

programmable multichip module design principles and applications. *Burman, Surendra, + , M-M Apr 93* 28-35

Programming; cf. Software design/development

Protection/safety

results of PRO-CHIP research on improving reliability of CMOS and MESFET ICs in safety-critical automotive electronics. *Zanoni, Enrico, + , M-M Feb 93* 30-48

Publishing; cf. Copyright protection; Printing

R

Radar; cf. Road vehicle radar

RD&E

HDTV research in Japan. *Kahaner, David K., M-M Oct 93* 49-53

research and development needs for advanced vehicle control systems as determined by Intelligent Vehicle Highway Society Committee. *Shladover, Steven E., M-M Feb 93* 11-19

Real-time systems

book review; µC/OS—The Real-Time Kernel (Labrosse, J. J.; 1992). *Mateosian, Richard, M-M Feb 93* 79

single chip for managing time-tagged priority queue in digitial audio processing. *Kahrs, Mark, M-M Aug 93* 49-51

Registers

architecture of Pentium microprocessor. *Alpert, Donald, + , M-M Jun 93* 11-21

shifting register windows to reduce spill/refill memory overhead. *Russell, Gordon, + , M-M Aug 93* 28-35

Resistance

thermal characteristics and thermal resistance of integrated circuits (On the Edge). *Bishop, Linda, M-M Aug 93* 52-58

Reviews; cf. Book reviews

Road transportation, automated highways

performance and architecture of high-dynamic-range camera chip for vision systems in traffic environments. *Seger, Ulrich, + , M-M Feb 93* 50-56

research and development needs for advanced vehicle control systems as determined by Intelligent Vehicle Highway Society Committee. *Shladover, Steven E., M-M Feb 93* 11-19

Road vehicle control

feedforward neural networks for lateral road vehicle guidance. *Neusser, Stefan, + , M-M Feb 93* 57-66

intelligent cruise control system that controls vehicle speed according to speed and distance of preceding vehicle and roadside information signals. *Palmquist, Ulf, M-M Feb 93* 20-28

Road vehicle control; cf. Road transportation, automated highways

Road vehicle electronics

automotive electronics (special issue). *M-M Feb 93* 7-28, 30-48, 50-66

interview with J. Rivard on ICs in automobiles (Micro View). *Myers, Ware, M-M Feb 93* 4-6

performance and architecture of high-dynamic-range camera chip for vision systems in traffic environments. *Seger, Ulrich, + , M-M Feb 93* 50-56

results of PRO-CHIP research on improving reliability of CMOS and MESFET ICs in safety-critical automotive electronics. *Zanoni, Enrico, + , M-M Feb 93* 30-48

Road vehicle radar

intelligent cruise control system that controls vehicle speed according to speed and distance of preceding vehicle and roadside information signals. *Palmquist, Ulf, M-M Feb 93* 20-28

Road vehicle reliability; cf. Road vehicle electronics

Robots

development of virtual reality systems in Japan; application to 3-D imaging of software structures and remote control of construction robots. *Kahaner, David K., M-M Apr 93* 66-73

Routing; cf. Multiprocessing, interconnection

S

Safety; cf. Protection/safety

Schottky FET logic circuits; cf. MESFET integrated circuits

Semiconductor device bonding; cf. Integrated circuit bonding

Semiconductor device mechanical factors; cf. Integrated circuit mechanical factors; Microelectromechanical devices

Semiconductor device packaging; cf. Integrated circuit packaging

Semiconductor device reliability; cf. Integrated circuit reliability

Semiconductor device thermal factors; cf. Integrated circuit thermal factors

Semiconductor memories

development of 3-D IC memories for high- and low-end systems. *Some, Rafael, +, M-M Dec 93* 6-7, 102

Semicustom integrated circuits; cf. Application-specific integrated circuits

Signal processing

single chip for managing time-tagged priority queue in digitial audio processing. *Kahrs, Mark, M-M Aug 93* 49-51

Signal processing; cf. Image processing

Social factors; cf. Technology social factors

Software debugging; cf. Software reliability

Software design/development

book review; Code Complete (McConnell, S.; 1993). *Mateosian, Richard, M-M Aug 93* 61-62

book review; Programming As If People Mattered (Broenstein, N. S.; 1991). *Mateosian, Richard, M-M Aug 93* 61

book review; Software Engineering: A Programming Approach, 2nd edn. (Bell, D., et al.; 1992). *Jordan, J. E., M-M Jun 93* 86

book review; The Elements of Friendly Software Design—The New Edition (Heckel, P; 1991). *Mateosian, Richard, M-M Feb 93* 79

book review; Writing Solid Code (Maguire, S.; 1993). *Mateosian, Richard, M-M Aug 93* 62

Richard H. Stern

Graham & James

2000 M Street, N.W.

Suite 700, Washington,

DC 20036-3307

Micro
Law

Patents and international trade issues

Two recent Washington, D.C., conferences on Japan-US trade and patent problems focused on a July 1993 report to Congress from the US General Accounting Office and what ought to be done about its conclusions. The report is titled "Intellectual Property Rights: US Companies' Patent Experiences in Japan."

The GAO report

At the request of US Senators Rockefeller (D-WV) and DeConcinci (D-AZ), the GAO surveyed US companies' experiences in obtaining patents in Japan, and the sources of US companies' problems with the Japanese patent system.

The GAO found that three times as many US companies responding to the survey were more dissatisfied with their experiences with the Japanese patent system (39 percent) than with the US (13 percent) or European (3 percent) patent systems. They said that Japanese patents take too long to issue (six to seven years vs. two years in the US), cost too much to prosecute, and have inadequate scope.

Two frequently cited complaints against the Japanese Patent Office's examination procedure were 1) patent application filing must be in Japanese, unlike the US practice (file patent application in any language, file translation later); and 2) the JPO does not hire enough patent examiners, which means it takes a long time to get around to examining a patent application. Difficulties in enforcing patents against infringers in Japan were also cited. These factors, it has been argued, make it unreasonably difficult for US firms to do business in Japan and protect their technological innovations.

JPA conference

On October 14, a delegation from the Japan Patent Association, an organization of Japanese corporations' patent agents, met with a group of US corporations' patent lawyers. The JPA delegation made two points. First, they pleaded for more understanding, and attributed many of the perceived problems to cultural barriers and to US lack of understanding of how things are done in Japan. Second, they said, "Nobody's perfect." That led into a discussion of problems that Japanese companies have with the way the US patent system works, which they would like us to fix. The JPA urged that curing patent imperfections should be considered "a two-way street"—both sides should endeavor to correct the imperfections of their respective patent systems.

Submarine patents. The main Japanese problem with the US patent systems was so-called submarine patents. These are not patents on underwater ships; they are "stealth" patents. They suddenly surface when nobody expects them, after pending for years in the darkness of the US Patent and Trademark Office (PTO), and they cover technology that everyone in an industry has been using for years. Since US patents run for 17 years from issue date, a submarine patent's 17-year monopoly may not even begin until a whole industry that it affects has adopted technology infringing the patent. That may occur even though nobody but the applicant knew that the patent was pending.

Examples are recent submarine patents on

- the basic idea of a microprocessor, that is, CPU, memory, and I/O, all on a single chip;
- using liquid crystal devices as light amplifiers by controlling their optical properties with a control signal;
- the laser;
- pull-down windows (menus);

> ### *The submariner makes no real technological contribution, but simply discredits the patent system.*

- hard drives that are 3.5 inches or smaller; and
- disk drive systems in which the head gets to about 200Å above the surface.

Sometimes, what is called a submarine patent is simply the result of mishaps in the PTO. Perhaps a patent application gets caught up in an interference with another patent application, and it takes years to resolve the issue of who is entitled to the patent. (This happened with the basic patent on polypropylene.)

Other times, and this is what riles many people, patentees keep filing slightly revised versions of their patent applications ("continuation-in-part" patent applications) one after another, so that the examinations go on for many years. During this lengthy period, applicants observe what is going on in the industry—in which direction the technology is moving. Then, they change their claims in one of their later patent applications to sweep up the technological advances that others have commercialized.

Possibly the most famous example of a submarine patent was George Selden's patent on the automobile (basically, on a four-wheeled vehicle driven by an internal-combustion engine). Selden filed an initial, rather con-

ceptual patent application before the US automobile industry existed. Then, he waited for others to develop the working automobile. Next, he caused his patent to issue. Finally, he used the patent to regiment the whole automobile industry into the "Automobile Trust" (except for Henry Ford, who fought the Trust and eventually broke the patent).

The JPA delegation suggested that the US should fix its patent system to stop the US PTO from issuing so many submarine patents. Since other countries do not allow them, the JPA felt, we ought to be able to do something.

Actually, many US companies do not like submarine patents, either. They perceive them as a perversion of the patent system—a form of extortion by a scheming type of patent applicant. The submariner makes no real technological contribution, but simply discredits the patent system by trying to rip off the actual innovators of an industry.

Those who oppose patent submarining have suggested a set of changes in the US patent law to eliminate future submarines:

- Make the term of US patents 20 years from the initial filing date, rather than the issue date. Then, if applicants keep changing the patent applications, they use up their own time.
- Publish any pending patent application after 18 months. Then, the industry that it affects will not be kept ignorant of what may someday issue as a patent.
- Put a cap on the length of time after the initial filing date during which pending claims can be broadened to sweep up advances in the industry (perhaps five years).

These three changes, if adopted, would probably eliminate the submarine patent problem in the US. While there is no widespread opposition to these measures, there is no ground swell of popular demand for them, ei-

ther. Moreover, other current legislative projects have a much higher national priority.

Form-factor patents. The JPA also expressed concern with the US PTO's willingness in the past to issue form-factor patents. These are patents on an old structural combination, with a dimensional limitation placed on some element. For example, suppose that the disk drive industry has been steadily moving to smaller disk diameters: 8-inch, 5.25-inch, 3.5-inch, ... Midway in this progression, X files a patent application on a disk drive identical in structure to the existing disk drive consisting of elements $A + B + C + D$, with the added proviso that element C shall be no bigger than 3.5 inches. That is the only difference. The US PTO has issued a number of patents of that type, to the vast displeasure of many or most in the industries affected by the patents.

The PTO now seems to be aware of the problem, however, and seems no longer to be issuing them. Still, the PTO may start issuing them again in the future. In any event, not only the JPA but many in the US dislike form-factor patents. (Incidentally, a form-factor patent can also be a submarine patent.)

Algorithm patents. Still another example of "nobody's perfect," according to the JPA delegation, is the US PTO's willingness at times to issue patents that preempt a new algorithm. Again, many in the US also oppose algorithm patents; they share the JPA group's concerns.

JIAP conference

On October 15, the Japan Information Access Project, a US university-backed effort, put on a conference with the evocative title "Patent Politics Between the US and Japan." Again, many of the same issues concerned the speakers.

The lead-off speaker was the US Commissioner of Patents and Assistant Secretary of Commerce, who spoke out against what he called "an insidious and parasitic disease." He asserted that Japan gets the cream of US technology,

but US companies cannot sell in Japan. The results are that wealth is drained out of the US and our national well-being is threatened. He focused on three major problems that US companies have with the Japanese patent system, problems that prevent US companies from being able to penetrate the Japanese market.

First, the JPO insists on having patent applications filed in Japanese—instead of allowing English filing with translation later (as with the US practice). The effect is to keep US companies from filing patent applications in Japan. According to the commissioner, the Japanese government has not given US officials any explanation for this policy, "so one can only speculate that the Japanese government has no desire to ease the burdens that foreigners face in attempting to obtain protection for their inventions in Japan."

Second, it takes the JPO at least five years to finish examining a patent, "because the JPO does not have a sufficient number of examiners to handle the volume of applications pending." While this delay may adversely affect Japanese patent applicants as well as US applicants, there is an important difference in the consequences, the commissioner said. Japanese firms are satisfied to cross-license one another, so that delay in issuance does not really make much practical difference. But US firms need "exclusive intellectual property rights"—rights to prevent others from using the new technology—to break into the new market until they have established themselves. He said that is "an edge that can be critical in an economy [Japan] where long-term relationships count a good deal more than small differences in price. The inability to obtain true exclusive rights in Japan blunts that competitive edge."

Why doesn't the JPO hire more examiners? According to the commissioner, Japan made a deliberate decision not to have enough patent examiners. The result is to hinder US

firms' entry into the Japanese market.

Third, after the JPO finishes examining a patent application, but before it issues any patent, the JPO allows anyone to start an opposition proceeding. If there are multiple opposers, each one has a separate opposition proceeding. Until all of them are completed, no patent issues. Other countries that permit opposition (and the US, in re-

---

We should learn from Bre'r Rabbit.

---

examination, which is similar to opposition) issue the patent first and then have a consolidated opposition proceeding. This is faster, and patent rights relate back to issuance if the patent is upheld, making delay for its own sake a less-effective tactic for potential opposers. In Japan the procedure is so time-consuming and costly, the commissioner said, you have to be an IBM to work under that kind of system.

All of these patent practices, the commissioner maintained, are part of a larger pattern designed to maintain barriers against foreign entry into Japanese markets. "The picture is no accident; it exists by design." The Clinton administration, he warned, is deeply concerned. It will not take a laissez-faire attitude, because of the serious threat posed to US well-being.

Another administration speaker, an intellectual property negotiator from the office of US Trade Representative, repeated the theme that there is not just one issue between US firms and the JPO, but rather a pattern of obstacles. She saw the root of it in a Japanese industrial paradigm for dissemination of technology, rather that giving innovators exclusivity (an exclusive right to make, sell, and use) to reward and encourage innovation.

When asked about the "two-way street" and whether the US would do anything about submarine patents or other problems that the Japanese perceive in the US patent system, her response was very macho and emphatically no! She explained that US trade and intellectual property negotiators do not regard the situation "as one of quid pro quo." Any imperfections in the US patent system, she said, are not a disincentive to technological innovation or obstacle to international trade. Therefore, they do not have to be traded off with the Japanese to get them to remove the imperfections of their patent system, which *are* disincentives of that kind. It's up to the Japanese to make the necessary changes in their system.

What's going on here?

I do not propose to question the merits of anyone's complaints and contentions. I assume that all of the claimed problems are real and that the complaints about them are well-founded. My question is one about style, posture, and tactics.

The US patent system has some serious imperfections, but we do not seem to be able to do anything about them because other matters (health care, for example) are more pressing. The Japanese say, "Nobody is perfect. Let's make this a two-way street, and both try to fix our systems." Does the US regard this as an opportunity, or just dismiss what the Japanese say as a mere put-on?

How is it an opportunity? US industry does not like US submarine, form-factor, and algorithm patents any more than the Japanese patent spokesmen do. Instead of telling the Japanese, "Stick it in your ear," as the administration's spokespersons are telling us they are doing, we should all learn from Bre'r Rabbit's response when he was threatened with the briar patch, "Please, Bre'r Fox, don't throw me into the briar patch!"

Our officials should tell Congress the following:

We can get the Japanese to agree to permit patent filing in English, hire more patent examiners, and change their opposition procedures. But there is a price. We have to agree to a quid pro quo. We have to stop issuing submarine patents, form-factor patents, and algorithm patents in the US.

It's a tough price, but US industry is willing to accept the burden, to bite the bullet, to get what it wants from the Japanese. Therefore, to give US industry what the GAO report and the administration say it needs from the Japanese patent system, we have to give them some changes that they want in our patent system. We and US industry are therefore willing to sacrifice future enjoyment of any more submarine, form-factor, and algorithm patents and change our law to abolish them.

That would be a win-win situation. But right now, it is just a lost opportunity, because our spokespersons are too short-sighted and too dedicated to macho posturing to take advantage of the chance to be "forced" to get rid of public nuisance.

Reader Interest Survey

Indicate your interest in this department by circling the appropriate number on the Reader Service Card.

Low 177   Medium 178   High 179
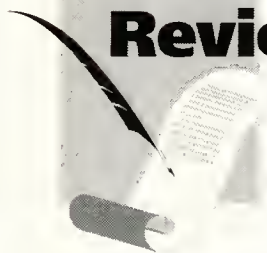
Richard Mateosian

2919 Forest Avenue

Berkeley, CA

94705-1310

(510) 540-7745

Desktop publishing tips plus Windows

The Underground Guide to Laser Printers, editors of *Flash Magazine* (Peachpit Press, Berkeley, Calif., 1993, 170 pp.; $12)

Flash Magazine provides nuts and bolts information about laser printers to a small cadre of loyal readers. This useful and practical book organizes and reprints some of *Flash Magazine's* most popular and helpful articles.

> Did you know that you could make printed circuit boards using your laser printer? Here's how:
>
> 1. Completely clean the copper surface with acetone or something that will remove any oils.
> 2. Pre-etch the board, or rough it up with a fine steel wool so it is a little rough and will grab the toner better.
> 3. Print your image in reverse (mirror image) on the label side of a full-page Avery label backing sheet, after having removed the adhesive label paper.
> 4. Preheat a T-shirt press up to 400 degrees F. We have also used a household iron with reasonable results. A dry-mount press might also work.
> 5. Place the circuit board in the heat press, copper side up, and the page on top of the copper, image side down.
> 6. Press it for 40 seconds to fully bond the toner to the copper.
> 7. Let it cool some, but not all the way, say, 30 to 60 seconds, and lift off the backing paper, leaving the toner on the copper.
> 8. Etch as usual in an acid bath.

I hope the authors will forgive the extended quote, but I can't think of a better way to convey the flavor of this Vermont-based publication. If hardware and software manuals were like this, perhaps everyone would use computers!

Other topics the book covers include how laser printers work, how to perform simple maintenance on all popular models, how to prevent excessive release of ozone by keeping your filters clean, how to buy a used laser printer, how to make full-color T-shirts with your laser printer (and special toners), and much, much more.

This inexpensive, concise book is one of the best bargains any laser printer owner will ever find. I hope they sell a million of them.

Graphics for the Desktop Publisher, Bruce T. Paddock (MIS Press, New York, 1993, 448 pp.; $29.95)

This is a book with many flaws, but its virtues outweigh them.

The best thing about the book is the way the author comes through. He is unmistakably a graphic artist, but with

a far above average grasp of what desk-top publishers need to know.

Paddock starts with a thorough survey of the kinds of hardware and software you might need to support graphics. His discussion of memory and storage requirements is especially helpful, because he understands the difficulties of storing and transporting the multitudes of large files that comprise graphics-heavy documents.

Next Paddock tells you the tricks of the trade, including how to use draw and paint programs, clip art, color, scanned images, and screen captures. He even tells you how to deal with service bureaus and print shops.

Paddock's weakest point is the way he handles the Mac-vs.-PC issue. He (or his publisher) wants a book that covers both, but he doesn't know the PC as well as he knows the Macintosh. He says the Mac provides a better environment for graphics and desktop publishing. He may be right, but I wish he were saying it from experience with both. He points out that many of the major Mac applications have Windows versions as well, and he tries to discuss applications generically. Still, when you look closely, you can see that he's talking about the Mac version.

The book is a little carelessly put together, with occasional typos, misnumbered figure references, and the like. On the other hand, it's attractive and nicely arranged, so it gives a good example of its subject matter.

Most of the book is in black and white with a little spot color—terms Paddock defines and elaborates on. Since he also needs to deal with full color, he has a small section of full-color figures on glossy paper. Paging through those figures, I found one of the book's subtler quirks.

In explaining the CMYK color system, Paddock points out that a mixture of cyan (C), magenta (M), and yellow (Y) in equal proportions is equivalent to black (K). He therefore displays two pinkish boxes and says: "The box on the left is filled with M

40%, Y 40%, K 10%. The box on the right is filled with C 10%, M 50%,Y 50%. There's no difference, is there?"

In fact, the numerical reasoning he used is an oversimplification, and the two CMYK mixtures (0,40,40,10) and (10,50,50,0) should be distinguishable. As I looked at them, they really appeared the same. I pulled out my Micronta 30X illuminated microscope (purchased on sale at Radio Shack for $5.95) and examined the two patches. Both were the (10,50,50,0) mixture. There was no (0,40,40,10) mixture. I hope that was really an accident. Maybe it's something to add to the section on how to deal with print shops.

If it hasn't come through yet, I should say that I really enjoyed this book, and I found a large part of it helpful and informative. I knew that I had found a kindred spirit when I read "Align is the greatest function ever included in a graphics application." Amen. If you don't learn anything else from this book, that alone is worth the price.

More Windows

Last issue (Oct. 93) I reviewed books from Microsoft Press covering Visual C++ and Visual Basic programming in the Windows environment. Other publishers have mined the same lode, and here are a couple of their nuggets.

The Windows Programming Puzzle Book, Kim Crouse (Wiley, N.Y., 1993, 461 pp.; $29.95)

The book bills its author, Kim Crouse, as America's premier Windows programming instructor. I can't judge that claim, but I like her book.

As other books in this format have proved, people enjoy the challenge of short puzzles. Crouse poses 101 puzzles, averaging just over a page each. She provides for each a short solution essay or sample program, averaging nearly three pages. You can read them in order or skip around. Each is independent and self-contained, but they are grouped into seven categories of related subjects.

Crouse's background as a Windows programmer, then a Windows instructor is crucial here. Many people could have made up 101 questions about Windows, but Crouse's arise from real situations or frequent errors and misunderstandings by students.

Crouse's puzzles aren't trivial questions about rare situations. They involve tasks that you can easily see a need for, like how to append data to the clipboard, how to display an appropriate picture for a minimized window as the user drags it, how to paint a gauge that fills with liquid.

Crouse's solutions are helpful. Rather than terse answers, she provides essays about the situation and its underlying principles and issues. Her sample programs are functional units. In fact, a companion disk (available separately for $20) contains code examples for about a third of the puzzles. I didn't receive a copy of the companion disk, so I can't speak from experience, but the advertisement in the back of the book claims that the disk contains makefiles or project files for various popular programming environments.

Windows programming is never going to be easy. Do yourself a favor by getting help from someone whose profession is giving help. You must enjoy challenges, or you wouldn't be programming Windows, so challenge yourself with these entertaining and informative puzzles.

The Visual Guide to Visual Basic for Windows, 2nd ed., Richard Mansfield (Ventana Press, Chapel Hill, N.C., 1993, 1,271 pp.; $29.95)

Let me confess up front that I haven't read all of this book. You won't either, but if you want to work with Visual Basic for Windows, you won't have to read much to get your money's worth.

Mansfield has put together an encyclopedia of Visual Basic. About 1,150 of the book's 1,271 pages are an alphabetically arranged set of essays. The 350 essays cover all of Visual Basics commands and functions. The rest of

the book consists of an overview and appendixes on the Windows application programming interface (API), debugging, operators, and companion products. One such product is a disk that Mansfield has prepared to accompany the book. I haven't seen it, so I can't comment on it.

Encyclopedias vary in the quality of their essays. A Britannica and a Funk and Wagnalls cover approximately the same topics, but no one would mistake one for the other. Mansfield's essays are closer to Britannica class. They are thorough, contain detailed examples, and discuss underlying principles intelligently.

Occasionally he wanders a little—I don't really need to know how the magnetron death tube clock works. Sometimes he is sickeningly gushy in praise of Visual Basic and Windows, as in "Windows is moving us toward the future of computing in ways more subtle and more powerful than simply its splendid visual features ... OLE may well be a major step on the road to artificial intelligence."

These are minor quibbles. This book is truly impressive. Its hard to believe that one person pulled it all together. Richard Mansfield and Ventana Press have a right to feel proud of this book. I doubt that you'll find a better book on Visual Basic.

Reader Interest Survey
Indicate your interest in this department by circling the appropriate number on the Reader Service Card.

Low 183   Medium 184   High 185

Dick Price

Staff Editor

Cows and computers

In many industries, a 1.9 percent annual rate of growth would be little cause for celebration. Returns so modest would send your typical money market manager hunting for new clients, if not a different career. Governments falter when the economy grows so slowly.

But in the US dairy industry, where recent Department of Agriculture figures show annual per cow increases in milk production approaching 2 percent, the story is different—especially when you consider that these rates reflect fundamental, sustained improvements to living animals.

"We point with considerable pride to that rate of genetic improvement," says George Wiggans, a research geneticist at the USDA's Animal Improvement Programs Laboratory in Beltsville, Maryland. His lab, part of the Agricultural Research Service, plays a pivotal role in the data collection and computer evaluation network fueling these dramatic increases. "In the 60s and 70s, the rate of increase in milk production was virtually flat. Starting in the late 70s and through the 80s, we saw a remarkable upsurge, which validates our evaluation programs."

Involved here besides the USDA is a close cooperation among dairy farmers, farmer-funded cooperatives and regional dairy records processing centers, breeder associations, and artificial insemination companies. Milk production records and data concerning body characteristics of dairy cows throughout the US are collected and sent to computerized processing centers, which maintain genealogical records and production histories. Wiggans' lab takes these data and calculates semiannual evaluations of the cows and bulls. Breeders, working through artificial insemination companies, use these evaluations to market bull semen of prize specimens to dairy farmers around the world.

Thanks largely to advances in computer technology that permit cheaper calculations with greater amounts of data, the frequency of evaluations can increase from twice yearly to nearly continuous. Such a development could accelerate the annual rate of genetic improvement by another 8 to 10 percent. "If milk production per cow is approximately 20,000 pounds per year," explains Paul Miller, an animal geneticist at American Breeders Service, "and if the rate of increase is 1.5 percent, another 10 percent increase would take us from 300 to 330 pounds per year."

Making the system work

To gather data that makes the system function, farmers participate in a program to keep track of each cow's milk production. Some also rate their herds for 14 body traits—udder width, foot angle, general stature—that reflect on a cow's suitability for breeding. Farmers also take monthly milk samples to determine protein and fat percentages. These data go to the processing centers, often now on computer disks or by modem.

The most advanced farms, using partly automated milking parlors, can collect milk production data electronically. An identifier implanted on the animal triggers sensors that record each cow's food and water intake, as well as milk production, automatically sending the information to a computer database. Such systems are rare. "Unfortunately, we haven't fully automated the protein sampling," says Wiggans. "Anyway, to meet our needs, an average of the production for the week gives us most of the benefits you would find with daily production records, without overwhelming us with data."

The cost to the farmer of the evaluation and breeding process is small. Indeed, breeding the herd costs less than 1 percent of total farm operating expenditures. "It's not a big ticket item,"

agrees Miller, though research tabs it as the single item giving the highest return on investment dollars. Nationwide, there are 10 million cows. "We're looking at a $200 million a year industry, plus insemination charges—so maybe $300 million."

Falling behind. Each state or region has its own farmer-funded management team to guide the data collection and reporting process. These cooperatives contract with the processing centers on a cost-plus basis to process the data and pass them along to the USDA. Currently, with PCs wresting away some of the mainframe's traditional functions, such centers find themselves in a bind. After beginning their computerization in the 50s as one of the original proving grounds for computer applications to industry, they now find themselves struggling to keep up.

From the data supplied to it, the USDA lab calculates and disseminates evaluations that rate bulls on the milk-producing performance and physical suitability of the cows they have sired. Breeders base semen purchases on these ratings. Ratings tend to fall, not because a bull's daughters produce less, but because, younger, genetically superior bulls supplant him at the top of the heap.

Typically, cows have a two-month rest period before calving each year, after which they start milking again. The cow's annual production—accumulated over a 10-month cycle—is called its lactation. "To speed up the process," says Wiggans, "we make predictions for the entire lactation based on a couple of months of information. We perform standardization processes on our system to provide an evaluation based on cows producing at the same time, taking into account genetic and environmental effects. The genetic effects are the goal of the process."

To keep pace, the USDA has also had to adapt to changing computer technology. In 1989, the current animal model system was introduced, which considers each cow's entire ancestry. Because of its large computing requirement, it was implemented using a National Science Foundation supercomputer at Cornell University. Acquisition last year of an IBM RISC System 6000 Powerserver to supplement its IBM 9370 Model 90 allowed Wiggan's lab to directly control all its computing resources. By sidestepping system changes beyond its control, the lab also avoids delays caused by moving tapes between computing centers and competing on shared machines. When the USDA evaluated workstations, memory considerations steered that choice. "Finding a workstation with 512 Mbytes of memory," says Wiggans. "That was the determining factor."

Herds changing worldwide. Progress in the US fuels dramatic improvement in dairy herds around the globe, with as much as 40 percent of America's semen sales and transfer embryos going overseas each year. "In Europe, they talk about the Holsteinization of the herds," says Miller. "That's the process of converting their Friesian cattle to Holsteins. Ironically, America's Holstein cattle were originally imported from the Friesland area of Germany and Holland. Now, after 60 years of selection for commercial profitability, our cattle are being used to upgrade the foundation stock of Europe."

Continuous evaluations

Now that the computer technology is in place, cries have gone up in some circles for more frequent, even continuous evaluations. More frequent reports, quicker turnaround, and more current data would speed genetic progress—by reducing the generation interval before elite specimens can be identified and used for breeding—and would also increase accuracy.

Who wins? Who loses? "As it is, dairy farmers wait with great anticipation for the genetic evaluations," says Denny Funk of the University of Wisconsin's Dairy Science Department. "They delay semen purchases until these new evaluations come out, then buy enough to last six months." More frequent evaluation would let them track ratings more closely, avoiding costly surprises. "That's like tracking the stock market," says Funk, who has also worked with the National Holstein Association, after the USDA the second largest player in this arena, which feeds its extensive pedigree records into the evaluation process. "Imagine if we could only look at the stock market twice a year."

"But dairy farmers are awfully busy," continues Funk. "They have a lot of management chores to do. Trying to monitor what's happening with individual bulls is maybe more than they want to deal with."

Artificial insemination companies, at least some of them, are dragging their feet as well. Accustomed to marketing their services with glossy brochures sporting color photographs of top-rated bulls, they complain that moving to more frequent evaluations would make such sales tools prohibitively expensive to print and too quickly out of date. Using on-line computer networks such as the Internet to disseminate information would fill the void, but some companies believe that customers would have unequal access to information and that marketing costs and confusion would increase.

Who pays? Once the USDA releases its evaluations into the public domain, private groups get no compensation for the information. And no matter how much of the load comput-

ers shoulder, continuous evaluations will mean more work for most of the players. "There's been some political structuring going on," argues Funk. "The USDA says the political aspect is not their problem—'we're here to analyze the data,' they say." Some industry segments refuse to provide free access to data they gather, not unless they get a cut of the action.

Working at a research-oriented, government-funded lab, Wiggans sees the USDA's role as making the process available, and then letting the other players decide how and how fast to proceed. "Historically, we've tried to develop a consensus," he says. "If we offer something, we want to make sure we can do it, which becomes more complex as we add traits to our evaluations. No doubt, we'll be able to provide better information and get it propagated more frequently—but only with industry cooperation."

Some would like things to move faster. "You can have a great idea," complains Miller, "but it can take years to get a consensus." He argues that the majority of farmers already favor the idea, citing surveys that show up to 70 percent of farmers backing more frequent evaluations. "The farmers want it. That's the reality."

More computers coming

Despite all the agricultural data gathering and evaluation, farmers themselves have been slow to catch onto the computer revolution. "We still have to wait for the infrastructure to develop," says Funk. "With the development of computer algorithms and supercomputers the technology is there for calculating the proofs more frequently. The biggest holdup is getting this information to the farmer. Not many dairy farmers have PCs yet, so downloading information more often would only reach a small group."

"Average herd size here in Wisconsin is still only about 55 cows," says Funk. "A herd that size is going to have a hard time justifying a computer. The next generation of dairy farmers, many of them are thinking 150, 200 cows, where they can capitalize on efficiencies of production. At that point, they're using computers for many aspects of their operations."

These advances in herd genetics and increased production may cut the number of farmers, making the trend toward computers doubly vital. "Americans have a love affair with cheap food," says Funk. "And we do have very affordable food here, so we'll probably increase production faster than consumption." Fewer cows will translate into fewer producers.

"If I wanted to survive in the dairy business," he concludes, "I'd want to position myself to produce milk as efficiently as possible. I'd be looking at management practices and systems that would assist me." Besides record keeping and business management computer applications tailored specifically for farming operations, dairy producers are beginning to use expert systems for analyzing masses of data.

To workstations

Besides these developments in the computer infrastructure, various innovations in processing and manipulating data have helped open the door to continuous evaluation of dairy cattle. Ignacy Misztal, a computer scientist turned animal researcher at the University of Illinois-Urbana, has been instrumental in these advances.

By developing and applying the matrix-free iteration process, he showed that PCs could handle the huge number of equations needed to evaluate all the nation's dairy herds. In these evaluations, each animal may result in several equations—remember the system contains 10 million Holstein cows. By partitioning equations into those that are global (needed all the time) and local (needed once during each round of an iteration), he was able to solve 7 million equations on a 386/33-MHz machine with 8 Mbytes of memory and a 600-Mbyte hard disk. "That approach was beneficial to all platforms," he reports, "but to PCs in particular. Many people realized then that PCs are not restricted to 'toy' problems."

Misztal chaired a symposium last spring in College Park, Maryland, held before the annual American Dairy Science Association meeting and designed to prepare the industry for the move to more frequent evaluations. He has also developed a demonstration program to help other nations with their evaluation programs. Called JAA and distributed free of charge, it is often the first program to run dairy evaluations, at least on a test basis, in many places. "I've counted 20 counties all over the world, including Russia and Australia, so far," says Mizstal. "I still get many requests."

The Holstein Association is already moving toward continuous evaluations, and Mizstal thinks that the USDA is sympathetic to the idea. With most of the regional processing centers having access to Unix workstations, all that's needed is industry agreement and installation of the communication links.

"The mainframe was the big roadblock," Mizstal contends. "It has little power and is very expensive. The one used by the Holstein Association works at about 7 MIPS, about the same speed as a 486/25-MHz machine for single programs. Any mainframe upgrade is in the hundreds of thousands of dollars, and common knowledge for programming mainframes in the network is very expensive. On the other hand, Unix power is very inexpensive, and networking is more or less built in. That makes Unix a good choice for compute-intensive applications like continuous evaluation."

Reader Interest Survey

Editorial comments

I liked: _____

_____

_____

I disliked: _____

_____

_____

I would like to see: _____

_____

_____

Reviewers Needed. *If interested, send professional data to Dante Del Corso, Dipartimento di Elettronica, Politecnico di Torino, C.so Duca degli Abruzzi, 24, 10129, Torino, Italy.*

For reader service inquiries, see other side.

PO Box is for reader service cards only.

IEEE Micro

PO BOX 16508
NORTH HOLLYWOOD CA 91615-6508
USA

IEEE COMPUTER SOCIETY

PO BOX 3014
LOS ALAMITOS CA 90720-9804
USA

New Products

Joe Hootman

University of
North Dakota

Design software/hardware

Comprehensive design environment

PIC Designer, a suite of integrated design tools, automates and shortens the cycle from design entry of programmable ICs to actual system implementation. When working with multiple PLD/CPLD/ FPGA architectures, users can evaluate the application of logic with a complete board design. This tool automatically selects devices, and partitions and fits a design to optimize performance based on user-defined parameters such as speed, cost, manufacturer, and logic family. Packaged with a 4,200-device library, it supports Synergy VHDL and Verilog synthesis, Verilog-XL and RapidSIM simulation, Allegro CBD layout, and Synopsys input through EDIF. *Cadence Design Systems, San Jose, CA, (408) 943-1234; from $16,000.*

Reader Service No. 10

Low-cost FPGA

Designed to provide customizable logic solutions for I/O and memory control functions in high-performance computing and communications systems, the iFX740 joins the FLEXlogic family of FPGAs. Using a segmented architecture with 40 macrocells arranged in four configurable function blocks, this FPGA has 2,500 usable logic gates or 5,128 bits of SRAM. Each CFB can be used as a block of 10-ns 24V10-like logic or 15-ns 128×10 SRAM or EPROM. Offering 80-MHz external clock frequencies, it consumes 1.0 mA per MHz. Packaged in 44- and 68-pin PLCC versions, the iFX740's segmented architecture provides a 10-ns propagation delay from any pin. *Intel, Santa Clara, CA, (602) 554-2388; from $17.90 for 44-pin PLCC version.*

Reader Service No. 11

6,000-gate FPGA

The A1460A, a 6,000-gate member of the ACT 3 family, offers 60-MHz overall system performance and 168 I/Os. Designed for graphics, imaging, high-speed PCs, workstations, and high-speed telecommunications, the device provides an 11.6-ns clock-to-out, pad-to-pad delay. Built using a 0.8-µm CMOS process, this FPGA comes in a 208-pin PQFP. *Actel, Sunnyvale, CA, (408) 739-1010; $195 (1,000s).*

Reader Service No. 12

DSP components

High-density floating-point modules

Based on AT&T's DSP3210 floating-point DSP, the Don Modules family of high-density floating-point modules comes in various processor and memory configurations. One, the 4/D, offers 133-Mflops peak performance and 64 Mbytes of DRAM. Measuring 3×3.3× 0.75 inches, the module provides a 32-bit nonmultiplexed data/address bus and four 25-Mbps serial ports with independent DMA. Compatible with both AT&T's VCOS and Apple's ARTA multimedia operating systems, it targets PC motherboards to high-end VMEbus CPU boards. The Don module ties each DSP3210 together via a common 32-bit local bus running at better than 45 Mbytes/s. *Ariel, Highland Park, NY, (908) 249-2900; Less than $1,000 (OEM quantities).*

Reader Service No. 13

SBus-based JTAG emulator

Designed for use with Sun Sparcstation and compatibles, the Brahma-JTAG and Brahma-MPSD systems provide either a JTAG interface for TI's TMS320C40 or an MPSD interface for TMS320C30 and TMS320C31 floating-point DSP

devices. With it, users can debug hardware and software embedded in a target system. A single Brahma SBus card and pod works for both C3x and C40 devices; only the target system cable must change to switch between interfaces. Extensions are available for Virtuoso, SPOX, and CL for Parallel C. *Sonitech International, Wellesley, MA, (617) 235-6824; $4,500 (JTAG or MPSD versions), $5,495 (combined version).*

Reader Service No. 14

Sonitech's Brahma

DSP, data acquisition board

Based on TI's TMS320C31 floating-point DSP operating at 33 MHz, the Model 310A DSP and data acquisition board provides up to 33-Mflops performance. Offering data acquisition for four 14-bit resolution differential channels with programmable gain and a 150-kHz sampling rate, it also includes one 12-bit, 300-kHz analog output. PC-to-Model 310A data transfer rates of 3 Mbytes/s are possible. Bundled software includes assembler, debugger, signal and spectrum display, and record and playback, plus the MODA program that manages multichannel data acquisition, simultaneous record and playback for stimulus/response applications, and advanced pretriggering options. *Dalanco Spry, Rochester, NY, (716) 473-3610; from $699.*

Reader Service No. 15

Low-power 32-bit DSPs

A low-power version of the TMS320-C31, the 32-bit TMS320LC31 achieves 33 Mflops and operates at a battery-saving 3.3 volts. Designed for lightweight, hand-held portability in devices such as bar code scanners, this DSP features two power-down modes to minimize power consumption. Also available is the 50-MHz TMS320C31. Delivering 50 Mflops, this version aids text-to-speech translation and enhanced 3D imaging. Housed in 132-pin PQFPs, these devices include an optimizing ANSI C compiler and high-level language debugger, XDS510 scan-based emulator, and PC half-card evaluation module. *Texas Instruments, Denver, CO, (713) 274-2320; $49 (TMS320-LC31), $81 (TMS320C31) (100s).*

Reader Service No. 16

Communications

PC/mainframe communications

Designed for easy installation and use, the NS/3270 emulation software package comes in 16-bit Windows and 32-bit Windows NT versions. Emulating IBM 3278 terminals (Models 2 through 5) and 3287 printers (DSC or SCS), the package includes two host sessions, file transfer, NetWare for SAA connectivity, and HLLAPI support. Capable of communicating over most LANs and WANs, NS/3270 operates in AdaptSNA LAN Gateway, Microsoft's SNA server for Windows NT, Novell's NetWare for SAA and NetWare SNA Gateway, and IBM's PC/3270 Gateway and OS/2 Communications Manager. *NetSoft, Laguna Hills, CA, (714) 768-4013; $95.*

Reader Service No. 17

Enterprisewide LAN

As an ESD system, Synchrony for the LAN provides automated software distribution management and retrieval on local and distributed LAN workstations. Synchrony delivers and installs software, applications, and updates; rolls back application changes; distributes data and reports; and collects data for consolidated reporting or backup. Designed to manage the deployment and maintenance of applications in a distributed computing environment, the package offers automated LAN capabilities for OS/2, Windows, and DOS users. A three-month trial program is available for $999. *Telepartner International, Farmington CT, (203) 674-2640.*

Readers Service No. 18

Small office, home office LAN

A peer-to-peer and client/serverlike LAN, Desk to Desk provides network access for 2 to 255 users. Sold with a complete office license, it includes standard LAN features: advanced printer sharing and print queuing and the ability to set up any PC as a workstation, server/station or dedicated server. Also featured are security provisions, compatibility with other LANs, DOS and Windows support, CD-ROM sharing, and laptop support. Full online documentation is included. *CBIS, Norcross, GA, (404) 446-1332; $129.*

Reader Service No. 19

Voice-processing DSP

Designed for the next-generation digital cellular market, the DSP56166 handles voice compression, modem, and control functions. This 16-bit general-purpose DSP features 60-MHz clock speeds, 30-MIPS performance, and a 33.3-ns instruction cycle. The device combines a high-speed core processor with on-chip sigma-delta codec, PLL, and OnCE on-chip emulation. Included on chip are 4K×16 data RAM internal memory, 256×16 program RAM, 4K×16 data ROM, and 8K×16 program ROM. *Motorola, Austin TX, (512) 891-2030; $60 (1,000s).*

Reader Service No. 20

Reader Interest Survey

Indicate your interest in the department by circling the appropriate number on the Reader Service Card

Low 189   Medium 190   High 191

Product Summary

Joe Hootman
University of North Dakota

Manufacturer	Model	Comments	RS #
Boards			
Mesa Electronics Emeryville, CA (415) 547-0837	4I27 PC/104 controller	LM629-based, two-axis DC servo motor control system on a stackable PC/104 bus card supports high-performance positioning systems. An 8-bit sign-magnitude PWM signal in each axis drives H-bridge servo amplifiers. Replacing a plug-in resistor network changes RC filter time constants. *$229 each (100s).*	80
Mesa Electronics	4C22 CPU	Optimized for low-cost, high-integration, instrument and controller applications is a CMOS, XT-compatible CPU implemented on the PC/104 bus. Features include a 2-Mbyte capacity, built-in disk emulator consisting of two 32-pin DIP sockets; a serial port, an LCD interface, and standard keyboard and membrane keyswitch interfaces. *$249 each (100s).*	81
Neuro DynamX Boulder, CO (303) 442-3539	XR25 accelerator	PC expansion card with i860 RISC, Dyna Mind software, and 2-Mbyte RAM trains neural networks for pattern recognition, signal classification, image processing, or financial forecasting in 4 minutes. Needs a 286/386/486 PC AT with ISA slot. *$3,495.*	82
Systems			
Motorola Tempe, AZ (602) 438-3287	Series 900 servers	Board-level modular, VMEbus/M88000-based servers and multiuser computers offer Posix-compliant Unix System V 4.0. Users can plug in peripherals and add expansion slots. Three modules are available: CPU plus VMEbus boards, general-purpose VMEbus capacity, and SCSI peripherals modules.	83
Sun Microsystems Mountain View, CA (415) 960-1300	UltraSparc-I CPUs	Advanced TI CMOS, 64-bit processor architecture provides 200 SPECint92 CPUs that conform to the Sparc International V.9 specification and execute four instructions simultaneously. An Early Access program available to licensees offers systems designers preproduction access to core technologies, development tools, and processor prototypes.	84
Software			
Compact Software Paterson, NJ (201) 881-1200	Circuit simulators	Release 6.0 for X-Windows of Super-Compact, Microwave Harmonica, and Scope CAD tools includes revised display managers that access linear and nonlinear tabular and graphical outputs. An enhanced tune mode, used together with multiple graphics updates, allows users to see circuit performance changes. *From $7,950 to $19,750.*	85
QTC Beaverton, OR (503) 626-3081	Numerical Advantage line	C or Fortran Math Advantage 5.0, Stat Advantage, and Spec Advantage libraries support supercomputers, mainframes, and workstations as well as Macintosh and PC platforms with 1,860 routines.	86

candidates for near-term commercial production are cache memory for high-performance CPUs, high-density SIMMs, and high-density memory boards where system form factors preclude the use of full stacks.

Cache memory. Current high-performance processors use several layers of cache memory in a memory speed hierarchy. Most of these machines use an on-chip, first-level cache followed by an off-chip, second-level cache. To achieve high-speed operation, manufacturers find it necessary to move to multichip modules (MCMs). A short stack of high-speed SRAMs brings the off-chip, second-level cache as close as possible to the CPU/cache controller, thus maximizing performance and reducing substrate costs. In an MCM the second-level cache resides in the processor package alongside the chip. Alternatively, a single-chip processor can have its secondary cache mounted directly on top of the die. This stack-on-processor configuration eliminates the primary on-chip cache, thereby reducing the processor IC cost by simplifying it and unifying the first- and second-level caches. Also, one configuration allows the processor chip set to include a separate cache controller IC. Mounting the short stack directly on top of the cache controller IC creates a single component, ready for packaging within an MCM or in a single-chip package. These "substrate-less MCMs" promise higher performance and lower cost.

High-density SIMMs and SIMM replacements. High-density SIMMs represent another area of interest. Ease of handling, cost, and I/O standards have made the SIMM the most widely used memory module. Portable systems will need more memory than can be obtained using current-generation SIMMs. If we replaced the single-chip memory packages used today with
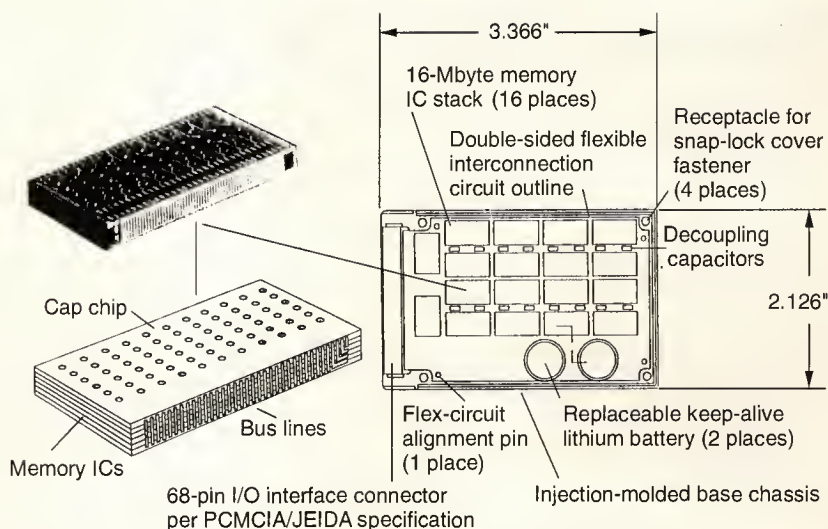
four- to eight-layer short stacks, the SIMM density would dramatically improve. This would lead to significant reduction in the total motherboard area dedicated to SIMMs. By including the decoder and driver functions on the SIMM, we make the short-stack SIMM appear to the system like any other SIMM, though it has a much higher memory capacity.

Memory boards, system upgrades. The computer system memory upgrade board represents a high-volume market for memory components. Both computer manufacturers and third-party vendors of compatible memory modules service this market. Designers use short stacks to upgrade board density by a factor of 4 to 16.

Another version of the add-in memory board is the PCMCIA or JEIDA (Japan Electronics Industry Development Association) memory card. In this application, for instance, designers could configure a memory card to provide up to 256 Mbytes of DRAM. A 68-pin Flash EEPROM card would provide up to 128 Mbytes of nonvolatile storage. These types of devices may, in the long run, obsolete rotating media

due to their ruggedness and low power consumption. Eliminating hard disks in laptop and palmtop computers, for instance, could extend battery life by up to an order of magnitude depending on the application. Figure 1 shows a 68-pin DRAM card using 8-layer short stacks to achieve 256 Mbytes in a credit card-size form factor.

For a variety of reasons, we expect 3D stacking to become far more cost-efficient in quantity over the long term than most of the MCM technologies now in place. As 3D technology continues to develop and evolve, and production problems are eliminated, the range of applications will broaden. Use of full and short stacks offer an appealing alternative to MCMs in an ever-increasing range of designs.

Figure 1. Plan view of the 256-Mbyte PCMCIA card, cover removed for clarity. (©1993 Irvine Sensor Corporation)

Labels in figure: 3.366" · 16-Mbyte memory IC stack (16 places) · Double-sided flexible interconnection circuit outline · Receptacle for snap-lock cover fastener (4 places) · Decoupling capacitors · 2.126" · Cap chip · Bus lines · Memory ICs · Flex-circuit alignment pin (1 place) · Replaceable keep-alive lithium battery (2 places) · 68-pin I/O interface connector per PCMCIA/JEIDA specification · Injection-molded base chassis

Reader Interest Survey

Indicate your interest in this department by circling the appropriate number on the Reader Service Card.

Low 192   Medium 193   High 194

IEEE MICRO Editorial Calendar

FEBRUARY 1994

Fault-tolerant systems
- Fault-injection techniques
- Treatment of transient faults
- Fault tolerance in parallel hardware systems
- Reconfiguration techniques

- **Portable electronics track**

Ad closing date: January 2

AUGUST 1994

Hardware-software codesign
- Hardware-software codesign, though perceived as an important and emerging field, presently offers few CAD tools for codesign. This issue highlights some recent achievements in these tools.
- **Portable electronics track**
- **Fault-tolerance track**

Ad closing date: July 1

APRIL 1994

Hot Chips V
- This extremely popular issue presents the latest developments in microprocessor and chip technology used to construct high-performance workstations and systems, as presented at the annual IEEE Computer Society TCMM-sponsored Hot Chips Symposium
- **Portable electronics track**
- **Fault-tolerance track**

Ad closing date: March 1

OCTOBER 1994

Multitheme issue
- Microsystem design education
- Microsensors
- Fuzzy system design
- Pen-based systems

- **Portable electronics track**
- **Fault-tolerance track**

Ad closing date: September 1

JUNE 1994

Analog VLSI and neural networks
- Architectures and circuits
- Cellular networks
- Pulse stream technique
- Signal processing
- Vision and sound
- **Portable electronics track**
- **Fault-tolerance track**

Ad closing date: May 1

DECEMBER 1994

Optical processing
- Design issues
- Communications
- Technology update

- **Portable electronics track**
- **Fault-tolerance track**

Advertising date: November 1

IEEE Micro helps designers and users of microprocessor and microcomputer systems explore the latest technologies to achieve business and research objectives. Feature articles in *IEEE Micro* reflect original works relating to the design, performance, or application of microprocessors and microcomputers. All manuscripts are subject to a peer-review process consistent with professional-level technical publications. *IEEE Micro* is a bimonthly publication of the IEEE Computer Society.

Articles may change. Please contact editor to confirm.

Intel chief gauges chip trends

Dick Price

Staff Editor

As chief architect for Intel's 386 and i486 microprocessors, John H. Crawford has his fingerprints all over the world's most prolific computer chip architectures. One of two design managers for the Pentium, he led a 100-plus-person company of Intel engineers through an exhaustive series of software simulations to fine tune its design. In his role as a leading proponent of the complex instruction-set architecture at the heart of Intel's x86 line, *Micro* approached Crawford for his views on the likely outcome of the ongoing architectural wars.

Isn't your job to dream up new trends for Intel to explore?

There's some truth in that, I guess. I do have a role in a future-generation microprocessor family, in addition to my responsibility for the 386 family. Historically, the trend in transistor growth from generation to generation has been a doubling in the number of transistors every two years, which creates terrific challenges.

But as you develop faster machines, don't you target fewer users who need that power?

That's an old argument. You forget there's something called the software spiral. Each new generation of computer can run the old software—better and faster. People want that instant response they get with a new design.

But then the new architecture spawns a whole new generation of software to take advantage of all the power that's suddenly available. Software designers see they can do exciting new things, and a whole new set of ideas springs up. The new software designs interact with the culture of computer users to create a whole new set of demands.

And that fuels new developments?

Right. Another thing is all the graphics capabilities that have become available, thanks to new generations of architectures. The future will give us more and more interactive communications among computer users because of these new developments.

Where does the Pentium fit?

The Pentium is designed for general-purpose computing. A large user base provides an established market for each new development. Pentium gives applications developers a big advantage—the latest generation of applications will run considerably faster on the Pentium, plus it will be used for network servers. It's designed for multiprocessing uses, as well.

Reports say Intel spends $900 million on research and development every year.

That was last year. And though the bulk of our R&D budget gets spent in developing semiconductor capabilities—a wonderful thing all around—there's all kinds of related development going on, too: packaging, computer-aided design, plus working on new generations of chips. For instance, we have recently been developing a broad spectrum of application-based benchmarks that will give a true picture of a computer architecture's performance.

Some say new RISC machines will compete against Intel's x86 line in a big way.

I hadn't noticed they needed any help. They're already a major factor in the marketplace. We see them competing for our customers, and our customers see them competing. RISC designs have made major inroads into enterprise computing, onto desktops, and for use as network servers. RISC designs have done wonderfully in the workstation market.